Cellular and Molecular Life Sciences (2023) 80:232
https://doi.org/10.1007/s00018-023-04812-w

ORIGINAL ARTICLE

BoGH13A$_{Sus}$ from *Bacteroides ovatus* represents a novel α-amylase used for *Bacteroides* starch breakdown in the human gut

Haley A. Brown[1] · Anna L. DeVeaux[1] · Brock R. Juliano[2] · Amanda L. Photenhauer[1] · Matthieu Boulinguiez[3] · Russell E. Bornschein[2] · Zdzislaw Wawrzak[4] · Brandon T. Ruotolo[2] · Nicolas Terrapon[3] · Nicole M. Koropatkin[1]

Received: 10 April 2023 / Revised: 15 May 2023 / Accepted: 17 May 2023 / Published online: 28 July 2023

Abstract

Members of the Bacteroidetes phylum in the human colon deploy an extensive number of proteins to capture and degrade polysaccharides. Operons devoted to glycan breakdown and uptake are termed polysaccharide utilization loci or PUL. The starch utilization system (Sus) is one such PUL and was initially described in *Bacteroides thetaiotaomicron* (Bt). BtSus is highly conserved across many species, except for its extracellular α-amylase, SusG. In this work, we show that the *Bacteroides ovatus* (Bo) extracellular α-amylase, BoGH13A$_{Sus}$, is distinguished from SusG in its evolutionary origin and its domain architecture and by being the most prevalent form in Bacteroidetes Sus. BoGH13A$_{Sus}$ is the founding member of both a novel subfamily in the glycoside hydrolase family 13, GH13_47, and a novel carbohydrate-binding module, CBM98. The BoGH13A$_{Sus}$ CBM98–CBM48–GH13_47 architecture differs from the CBM58 embedded within the GH13_36 of SusG. These domains adopt a distinct spatial orientation and invoke a different association with the outer membrane. The BoCBM98 binding site is required for Bo growth on polysaccharides and optimal enzymatic degradation thereof. Finally, the BoGH13A$_{Sus}$ structure features bound Ca^{2+} and Mn^{2+} ions, the latter of which is novel for an α-amylase. Little is known about the impact of Mn^{2+} on gut bacterial function, much less on polysaccharide consumption, but Mn^{2+} addition to Bt expressing BoGH13A$_{Sus}$ specifically enhances growth on starch. Further understanding of bacterial starch degradation signatures will enable more tailored prebiotic and pharmaceutical approaches that increase starch flux to the gut.

Keywords Microbiota · Alpha-amylase · Starch · Bacteroidetes · Carbohydrate-binding module · Glycoside hydrolase family 13

Abbreviations

AP	Potato amylopectin
Bo	*Bacteroides ovatus*
Bt	*Bacteroides thetaiotaomicron*
CBM	Carbohydrate-binding module
G2	Maltose
G3	Maltotriose
G4	Maltotetraose
G5	Maltopentaose
G6	Maltohexaose
G7	Maltoheptaose
GH	Glycoside hydrolase
Glc	Glucose
PS	Potato starch
PUL	Polysaccharide utilization loci
Sus	Starch utilization system

✉ Haley A. Brown
haleybr@umich.edu

✉ Nicole M. Koropatkin
nkoropat@umich.edu

[1] Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[2] Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

[3] Architecture et Fonction des Macromolécules Biologiques, UMR 7257, CNRS AMU; USC1408 INRAE, 13288 Marseille, France

[4] Synchrotron Research Center, Life Science Collaborative Access Team, Northwestern University, Lemont, IL, USA

Introduction

The gut microbiota is a consortium of bacteria, viruses, archaea and fungi in the large intestine with a vast reservoir of metabolic potential beyond that of the host [1–4]. Bacteria interact intricately with colonic tissue and can therefore

dictate health and disease outcomes in certain diet and host backgrounds [5]. Some interactions are direct, in the case of bacteria consuming mucus and colonizing the mucosal surface [6]. Conversely, relationships can be more indirect in the form of chemical signals that influence host energy storage, immune development and cell proliferation [7].

One example is in bacterial processing of dietary fiber, carbohydrate that traverses the gastrointestinal tract unmodified by host enzymes [8]. Estimates suggest that as much as 10% of calories can be harvested from the diet via bacterial carbohydrate fermentation [9]. Polysaccharide fermentation products in the form of short-chain fatty acids (SCFAs) have numerous beneficial impacts on host health [7]. Some bacteria are considered carbohydrate specialists, only encoding some glycoside hydrolases (GHs) to digest a few substrates [10]. However, many species in the abundant Bacteroidetes phylum, including *Bacteroides thetaiotaomicron* (Bt), deploy an extensive arsenal of GHs and polysaccharide lyases (PLs) to metabolize a staggering number of polysaccharides [11–15].

Dietary fiber degradation in Bacteroidetes is accomplished via the coordination of proteins encoded in polysaccharide utilization loci, or PUL [15–17]. The discovery of the PUL that target starch in Bt, the starch utilization system (Sus), characterized by Abigail Salyers and colleagues [18, 19] (Fig. 1A) has served as a model for the

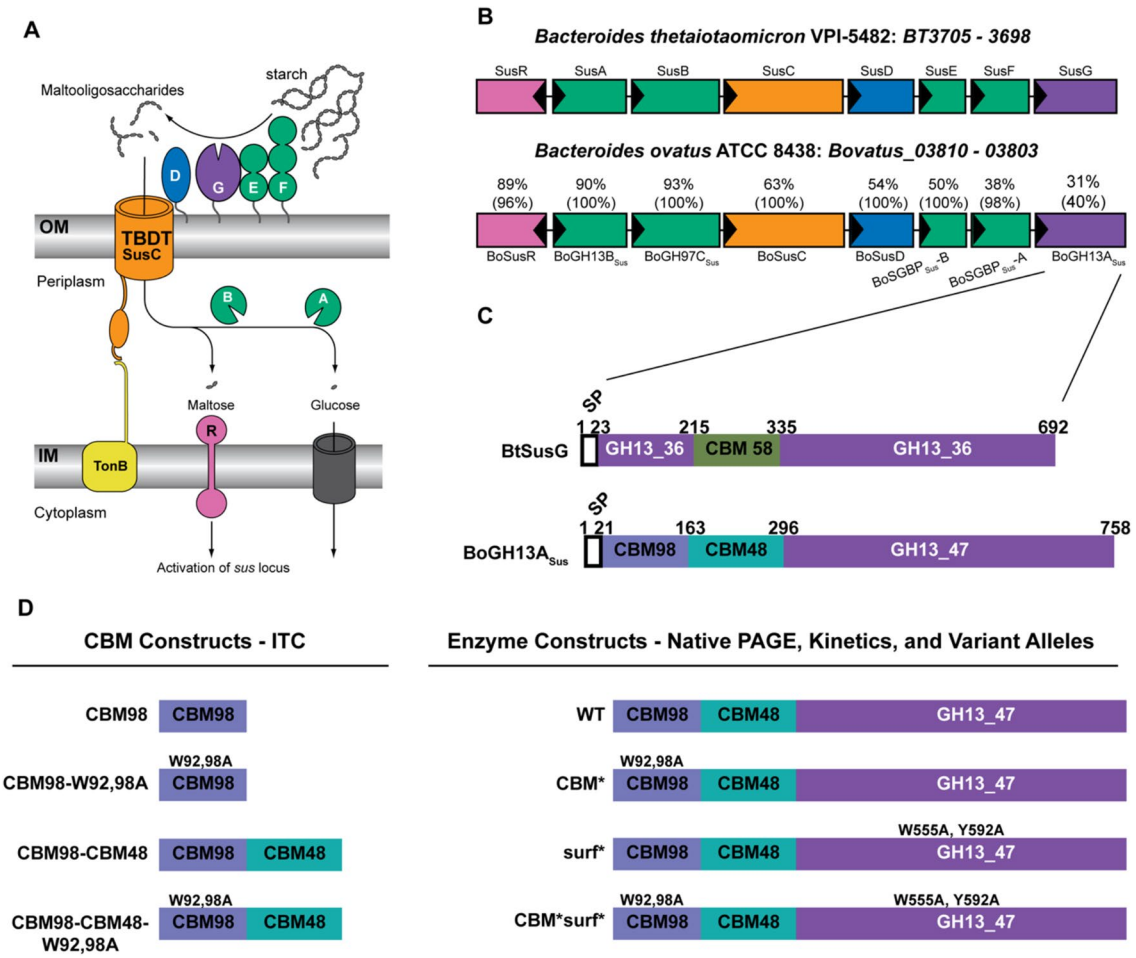


Fig. 1 Bt and Bo Starch Utilization Systems and extracellular α-amylase proteins. **A** Cartoon representation of the starch utilization system (Sus) from Bt. SusD-G are outer membrane lipoproteins. SusC is a TonB-dependent transporter. Periplasmic SusA and SusB complete starch breakdown into glucose. Maltose released during saccharification binds the SusR regulator to increase expression of *susA-G*. **B** Overview of *sus* loci in Bt and Bo. Locus tags *BT3705–3698* correspond to SusR-G. *Bovatus_03810–03083* correspond to BoSusR–BoGH13A$_{Sus}$. Genes are colored the same as their representation in (**A**). The percentages above the Bo genes indicate the amino acid sequence identity, and the coverage is given in parentheses. For example, BoSGBP$_{Sus}$-B is 38% identical to SusE over 98% of the SusE sequence. Genes are drawn to scale. Adapted from Foley et al. [20]. **C** Schematic of the SusG and BoGH13A$_{Sus}$ primary amino acid sequences, at scale. A CBM58 domain interrupts the GH13-coding sequence in SusG. BoGH13A$_{Sus}$ contains a CBM98 and CBM48 at its N terminus. The amino acid numbers delineate the domain boundaries. **D** An overview of constructs referred to throughout the text. Note that the signal peptide was not included for recombinant protein expression but is included for variant allele production in *Bacteroides*

molecular, biochemical and structural understanding of the PUL concept. In most PUL, a single extracellular enzyme cleaves polysaccharides for import and depolymerization away from competing organisms [17]. One or more surface glycan-binding proteins (SGBPs) may or may not be present. But, a SusC/D pair, additional surface exposed or periplasmic enzymes and an ability to sense oligosaccharides via a regulator are ubiquitous PUL features [15].

In BtSus, starch is degraded at the cell surface by an endo-acting GH13 amylase lipoprotein, SusG which is required for growth on starch [21–24]. In the absence of SusG, Bt can grow on maltooligosaccharides up to DP40 [25]. SusG is mobile on the outer membrane, only forming a complex with stationary SGBPs SusE and SusF in the presence of starch [26–28].

The SusG catalytic domain falls into subfamily 36 of GH family 13 according to the Carbohydrate Active Enzyme database classification (CAZy.org; [29]). It adopts a classic $(\beta/\alpha)_8$ TIM barrel fold that is interrupted within the B domain, located between β3 and α3, by a carbohydrate-binding module from family 58 (CBM58) that binds starch [23]. The placement of this CBM is unusual, as CBMs are more frequently appended to the N or C termini of GHs, with some notable exceptions including predicted starch-binding CBM20- and CBM21-containing proteins and some GH10 xylanases [30–32]. In addition to the CBM58, the SusG catalytic domain has a surface starch-binding site adjacent to the active site. The CBM58 is required for full activity of the enzyme on insoluble starch, but not on soluble polysaccharide, while the mutation of the surface site results in a mild decrease in activity on soluble starch and a moderate decrease on insoluble starch [23].

Starch PUL analogous to BtSus are among the most frequent in the Bacteroidetes phylum and are widespread in the Bacteroidia class that mostly consists of gut constituents [33]. The goal of this work was to investigate the structural and functional differences between SusG and its analog from *Bacteroides ovatus,* BoGH13A$_{Sus}$. These enzymes display extremely low sequence similarity despite very high synteny and conservation of the other Sus components (Fig. 1B). We performed structural studies and in vitro binding assays to characterize a new CBM family in BoGH13A$_{Sus}$, CBM98. We also observed the contribution of CBM98 to Bo starch uptake in vivo and how this compares to SusG. Furthermore, substrate preferences in BoGH13A$_{Sus}$ were analyzed since its catalytic domain represents a novel GH13 subfamily, GH13_47, distinct from that of SusG, with a unique preference in α-amylases for binding Mn^{2+}. Our data underscore the importance of individual PUL analysis for designing dietary interventions that influence discrete species toward improved health outcomes.

Results

BoGH13A$_{Sus}$ has a distinct evolutionary origin from SusG

The closest homologs to SusG can be found in Bt and *Bacteroides faecis* strains (> 97–92% sequence identity) with the next closest relatives in *Phocaeicola dorei* strains (67% identity) [29]. However, in many Bacteroidetes species, the starch PUL show higher conservation levels of non-extracellular α-amylase components while using a different, but possibly analogous, α-amylase. *Bacteroides ovatus* (Bo) is one such organism and encodes an extracellular α-amylase, BoGH13A$_{Sus}$, that demonstrates a huge sequence discrepancy compared to SusG with only 40% coverage and 31% sequence identity. On the other hand, the Bo homologs of the SusR regulator and SusAB periplasmic enzymes exhibit 88–93% identity and homologs of surface SusCDEF proteins have good conservation levels to the equivalent Bt proteins (Fig. 1B). BoSusD, BoSGBP$_{Sus}$-B and BoSGBP$_{Sus}$-A are predicted to have analogous starch-binding residues to those in SusDEF, respectively (Supplementary Fig. 1).

BoGH13A$_{Sus}$ contains an N-terminal CBM-like module that defines a new family, CBM98, and an additional starch-specific CBM48, while it lacks the CBM58 embedded within the GH13. Further, the BoGH13A$_{Sus}$ GH13 catalytic module does not belong to the GH13_36 subfamily but defines a novel GH13 subfamily, GH13_47 (Fig. 1C; discussed in more detail hereafter [29]).

This demonstrates that despite the microsynteny between Bt and Bo Sus genomic regions, BoGH13A$_{Sus}$ and SusG have a distinct evolutionary origin, the functional consequences of which have yet to be assessed.

BoGH13A$_{Sus}$ has an enzymatic profile distinct from SusG

The BoGH13A$_{Sus}$ GH13 domain, comprised of prototypical α-amylase domains A, B and C, is approximately 60 residues shorter than that of SusG and the two GH13s align with just 21% sequence identity (Supplementary Fig. 2A,B) [34].

Like most amylases, BoGH13A$_{Sus}$ does not break down the cyclic oligosaccharides α- and β-cyclodextrin (αCD and βCD, respectively), whereas SusG can break these down, albeit not efficiently [23]. The smallest maltooligosaccharide that BoGH13A$_{Sus}$ hydrolyzes is maltotetraose (G4), while SusG can hydrolyze maltotriose (G3) [23] (Fig. 2). The fact that glucose is not produced from G4 hydrolysis implies that four subsites in the BoGH13A$_{Sus}$

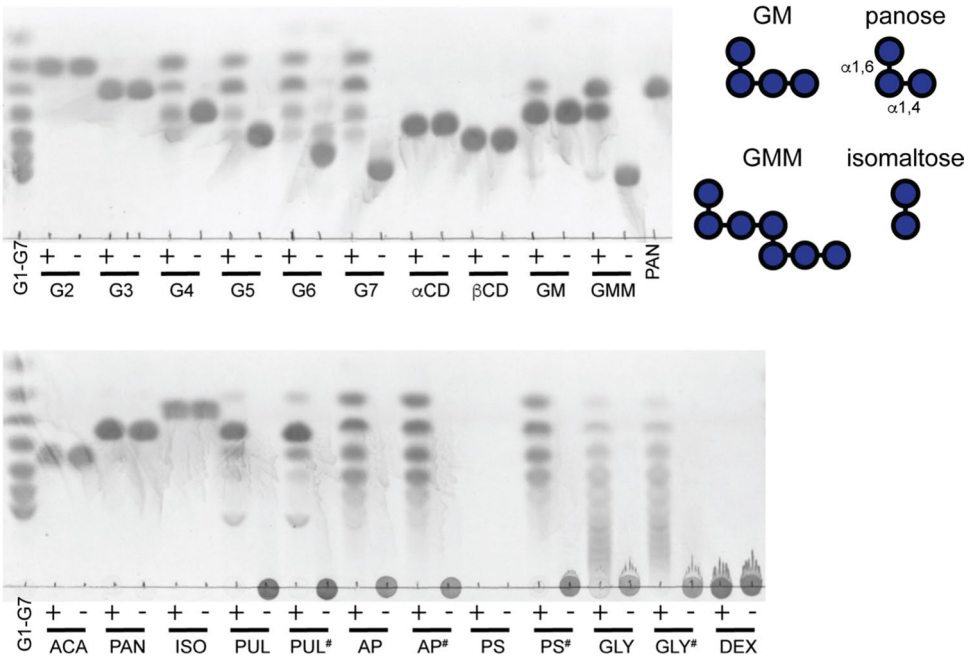


Fig. 2 BoGH13A$_{Sus}$ releases maltose and larger oligosaccharides. 0.5 µM BoGH13A$_{Sus}$ was incubated at 37 °C overnight in 20 mM HEPES, 100 mM NaCl, pH 7.0 with 5 mg/ml oligosaccharide or polysaccharide. 2.5 µl of each sample was loaded onto a TLC plate. Standards and controls lacking BoGH13A$_{Sus}$ are included. G1—glucose, G2—maltose, G3—maltotriose, G4—maltotetraose, G5—maltopentaose, G6—maltohexaose, G7—maltoheptaose, αCD—α-cyclodextrin, βCD—β-cyclodextrin, GM—6^3-α-D-glucosyl-maltotriose, GMM—6^3-α-D-glucosyl-maltotriosyl-maltotriose, PAN—panose, ACA—acarbose, ISO—isomaltose, PUL—pullulan, AP—potato amylopectin, PS—potato starch, GLY—bovine liver glycogen, DEX—dextran. #Indicates autoclaved carbohydrate; + indicates BoGH13A$_{Sus}$ was added; − indicates BoGH13A$_{Sus}$ was omitted. The inset contains representations of GM, GMM, PAN and ISO to show the position of α1,6 bonds referred to in the main text

active site, from - 2 to + 2, must be occupied for catalysis to occur. The production of G3 and absence of glucose from G4 hydrolysis are likely due to transglycosylation; during the reaction, maltohexaose (G6) is generated via transglycosylation of G4 and G2 and subsequently broken down to G3 which cannot serve as a substrate. Transglycosylation of two G4 molecules to G8 could subsequently be hydrolyzed to G3 and G5, which would explain the formation of the latter. (Fig. 2).

Like SusG, BoGH13A$_{Sus}$ can tolerate α1,6 bonds and breaks down 6^3-α-D-glucosyl-maltotriosyl-maltotriose (GMM) into panose and 6^3-α-D-glucosyl-maltotriose (GM) [35] (Fig. 2). The α1,6-linkages would span the − 3/− 2 and + 1/+ 2 subsites in this case [36]. These data support the results observed in pullulan hydrolysis where the predominant product is panose. A small amount of panose and glucose is produced from GM which is surprising given glucose is undetectable from the breakdown of α1,4-containing oligosaccharides (Fig. 2). The α1,6 bond at the non-reducing end of GM may help anchor the substrate such that the + 2 subsite does not need to be occupied, unlike what we observed with G4. Structural evidence below details how BoGH13A$_{Sus}$ accommodates an α1,6 bond at the − 2 position, which is analogous to how GM would be accommodated. BoGH13A$_{Sus}$ does not target α1,6 bonds as the enzyme does not break down panose, isomaltose or dextran (Fig. 2).

The BoGH13A$_{Sus}$ product profile from the hydrolysis of polysaccharides is the same as that found in SusG except that, as observed for the oligosaccharides, glucose is not released [23]. BoGH13A$_{Sus}$ breaks down all autoclaved polysaccharides tested (indicated by a$^{\#}$) but does not break down granular potato starch. Compared to amylopectin, BoGH13A$_{Sus}$ leaves behind larger oligosaccharides during glycogen hydrolysis (Fig. 2). This may be due to the presence of more α1,6 branch points in the latter although BoGH13A$_{Sus}$ tolerates the single α1,6-linked glucose at the non-reducing end of GM and GMM [37].

The BoGH13A$_{Sus}$ structure features an opposite GH orientation and CBM position relative to the cell membrane compared to SusG

The 1.9 Å resolution structure of BoGH13A$_{Sus}$ (residues 22–758) was solved by molecular replacement in the space group P2$_1$ with four molecules in the asymmetric unit (R_{work}/R_{free} of 20.9%/27.0%, Table 1). All four protein monomers align with a root-mean-square deviation (rmsd)

Table 1 BoGH13A$_{Sus}$ crystallographic data and refinement statistics

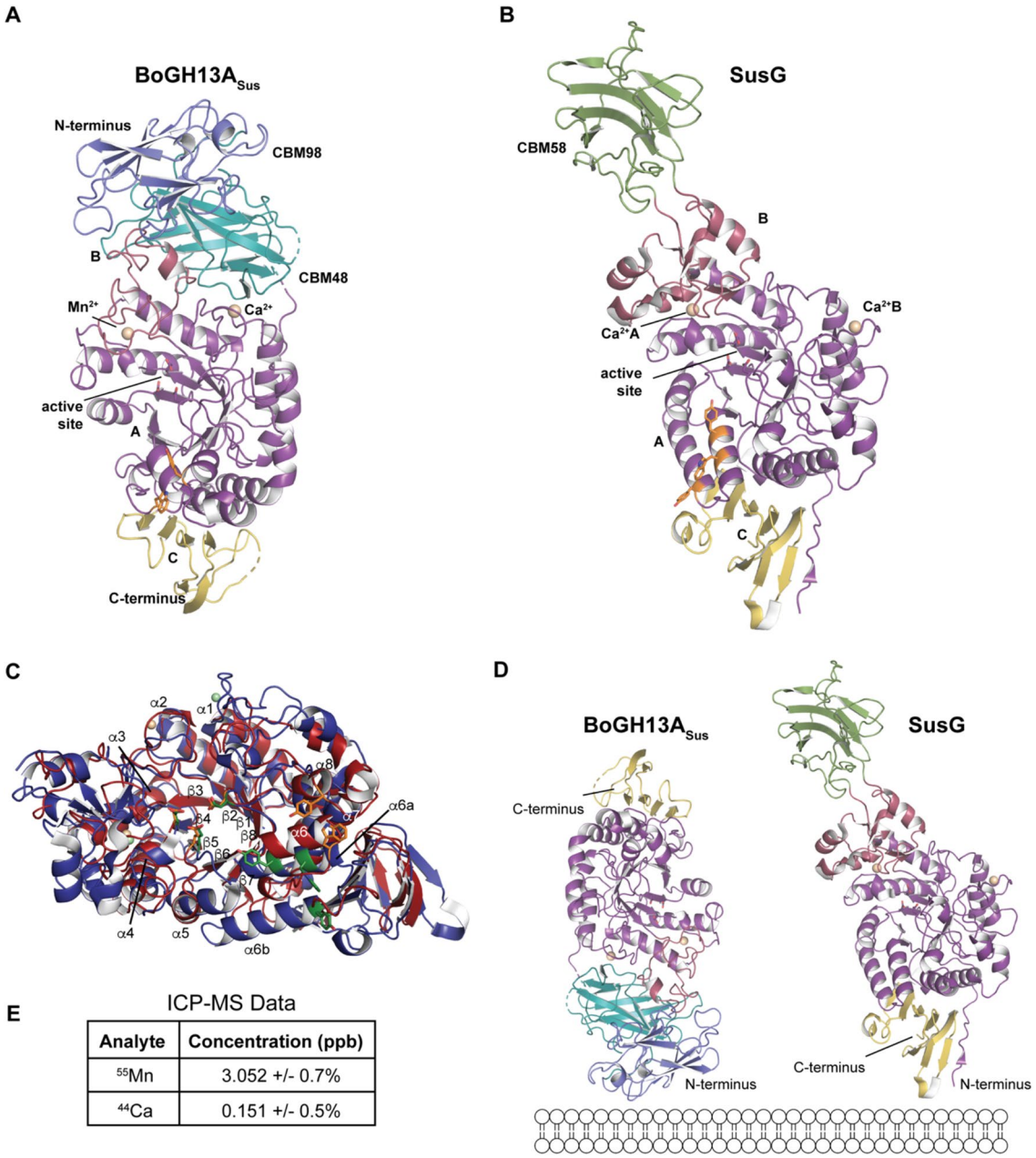
	WT—native	E523Q—maltoheptaose	WT—acarbose
PDB code	8DGE	8DL1	8DL2
Resolution range (Å)	62.0–1.89 (1.96–1.89)	34.54–2.09 (2.17–2.09)	38.6–1.99 (2.06–1.99)
Space group	$P2_1$	$P2_1$	$P2_1$
Unit cell	*a* = 100.39 Å *b* = 148.44 Å *c* = 112.77 Å β = 91.0°	*a* = 99.68 Å *b* = 128.64 Å *c* = 149.76 Å β = 105.37°	*a* = 100.08 Å *b* = 125.31 Å *c* = 150.78 Å β = 102.1°
Total reflections	1,323,840 (120,261)	838,527 (63,708)	1,145,170 (115,127)
Unique reflections	262,849 (25,868)	212,239 (21,205)	244,819 (23,974)
Multiplicity	5.0 (4.6)	3.9 (4.1)	4.7 (4.8)
Completeness (%)	99.7 (98.5)	98.7(99.2)	98.5 (96.9)
Mean *I/σ (I)*	8.8 (1.5)	10 (1.2)	9.6 (1.3)
Wilson *B*-factor (Å^2)	27.02	20.84	19.83
R-merge (%)	9.5 (95.4)	7.4 (112.3)	8.6 (104.8)
CC1/2	0.997 (0.688)	0.998 (0.467)	0.997 (0.587)
CC*	0.999 (0.903)	0.999 (0.780)	0.999 (0.86)
Reflections used in refinement	262,422 (17,911)	212,221 (21,206)	244,780 (23,973)
Reflections used for R-free	13,119 (942)	10,581 (989)	12,244 (1265)
R-work %	20.9 (26.1)	17.7 (30.1)	17.5 (30.8)
R-free %	27.1 (35.7)	22.3 (33.1)	21.9 (31.6)
TLS Groups	4	/	/
Number of non-hydrogen atoms	24,832	25,543	25,774
Macromolecules	22,790	22,701	22,792
Ions	8	10	9
Ligands	470	945	697
Water	1,564	1,887	2,276
Protein residues	2,842	2,839	2,845
RMS (bonds, Å)	0.013	0.017	0.014
RMS (angles, °)	1.71	1.93	1.87
Ramachandran favored (%)	96.05	95.13	95.51
Ramachandran allowed (%)	3.7	4.65	4.28
Ramachandran outliers (%)	0.25	0.21	0.21
Clashscore	2.5	3.8	2.38
Avg. B factor (Å^2)	31.5	28.4	26.4
Macromolecules (Å^2)	31.0	27.7	25.7
Ions (Å^2)	30.1	34.3	25.4
Ligands (Å^2)	40.5	39.3	32.4
Solvent (Å^2)	35.2	32.6	31.7

Highest resolution shell shown in parentheses

of ~ 0.3 Å. The first ~ 20 amino acids in each molecule were not resolved which likely reflects a flexible linker that connects the first domain of the protein to the outer membrane via a thioester linkage at a Cys proceeding the signal peptide (SPII). From the N to C terminus, BoGH13A$_{Sus}$ is comprised of two CBMs followed by a canonical amylase fold with A, B and C domains as typified by this glycoside hydrolase family (Fig. 3A).

BoGH13A$_{Sus}$ is compact, with the two CBMs packed closely together and next to the catalytic domain. There is ~ 2110 Å^2 of buried surface area in the CBM98/CBM48 + GH13 interface and ~ 2600 Å^2 of buried surface area between the CBM48 and GH13 domains. The interface between CBM98 and CBM48, along with a small portion of the GH13 B domain, is held in place by several hydrogen-bonding interactions (Supplementary Fig. 3A–C). Residues A90 and N91, which precede the CBM98-binding platform of W92/W98 discussed below, hydrogen bond to CBM48 via T283 and Q286, which may help orient and stabilize this binding site (Supplementary Fig. 3A).

Within the CBM48/GH13 interface, we were surprised to observe that E274 of CBM48 coordinates the structural Ca^{2+}



E

ICP-MS Data

Analyte	Concentration (ppb)
^{55}Mn	3.052 +/- 0.7%
^{44}Ca	0.151 +/- 0.5%

ion contained in the A domain (Supplementary Fig. 3D). Typically, a Ca^{2+} ion at this position in the A domain is held in place by residues only from that domain [23, 40–42]. The extensive interface between the CBM48 and catalytic domain, including the shared Ca^{2+}, is likely why attempts to express and purify the GH13 domain alone were unsuccessful. Moreover, it is likely that BoGH13A$_{Sus}$ remains a compact structure without significant flexibility or alternative conformations between the domains in solution.

A DALI search [43] with the catalytic domain of BoGH13A$_{Sus}$ reveals structural similarity with α-amylases from *Anoxybacillus* sp. SK3-4 (*Z*-score of 35.2) and *Geobacillus thermoleovorans* (*Z*-score of 35.0) [44, 45]. These proteins are part of a recently assigned GH13_45 subfamily based on [46] of which BoGH13A$_{Sus}$ is not a part. The most striking differences between BoGH13A$_{Sus}$ and these homologs reside in the B domains which is typically involved in shaping substrate specificity (Supplementary Fig. 3E).

To compare the three-dimensional structures of SusG and BoGH13A$_{Sus}$, we aligned their GH13 domain and separated the proteins for clarity in Fig. 3A (BoGH13A$_{Sus}$) and 3B (SusG). SusG lacks the two N-terminal CBMs seen in BoGH13A$_{Sus}$ and instead has a CBM58 that protrudes from the B domain (Fig. 3B) [23]. The CBM58 does not make any hydrogen-bonding contacts with the GH13 domain and is

◂**Fig. 3** Structural comparison between BoGH13A$_{Sus}$ and SusG. **A** Ribbon representation of the native BoGH13A$_{Sus}$ structure. CBM98 in BoGH13A$_{Sus}$ is colored in slate and corresponds to residues 43–163. CBM48 is colored in teal and corresponds to residues 164–296. The A domain is colored in purple, B domain in burgundy and C domain in yellow. The A domain is comprised of residues 297–403 and 451–680; the B domain, residues 404–450; and the C domain, residues 681–758. Ca^{2+} or Mn^{2+} ions are colored in wheat and represented as spheres. Surface binding site residues in **A** and **B** are shown as orange sticks. The 27 amino acid linker between CBM98 and CBM48 is nearly complete in chains **A**–**C**, while the start and finish of the 17 amino acid linker between CBM48 and the catalytic domain are unresolved in each chain. Finally, a solvent exposed loop in the C domain centered at N733 and the C terminus is not well resolved in all chains. **B** The previously published native structure of SusG is presented (PDB ID: 3K8K, [23]). SusG amylase domains are colored as in panel A. The CBM58 is colored in green. To show the proteins in A and B, the GH13 domains were aligned and separated into different panels. **C** Ribbon representation looking down the A domain barrels of BoGH13A$_{Sus}$ and SusG. The entire GH13 of each protein is shown but CBMs were omitted for clarity. BoGH13A$_{Sus}$ is colored in red and SusG in blue. α-Helices and β-strands are labeled from N to C terminus. SusG has two α6 helices labeled a and b. BoGH13A$_{Sus}$ metals are colored in wheat, and those from SusG are colored light green. The catalytic triad and surface binding site residues (detailed more in Fig. 4) are shown as orange sticks in BoGH13A$_{Sus}$ and green sticks in SusG. The A domains were aligned using the "super" command in PyMol. **D** BoGH13A$_{Sus}$ (left) and SusG (right) are displayed based on their predicted orientation with respect to the outer membrane, as both have an N-terminal lipid embedded in the membrane. A-D were rendered in PyMol [38, 39]. **E** The concentrations of ^{55}Mn and ^{44}Ca isotopes BoGH13A$_{Sus}$ as calculated by inductively coupled plasma mass spectrometry (ICP-MS) quantitation data. Calibration curves were constructed for both ^{55}Mn and ^{44}Ca to calculate the concentrations of these species in BoGH13A$_{Sus}$. The ^{44}Ca isotope was used instead of ^{40}Ca to minimize interference from argon in the plasma

positioned 40 Å away from the active site. This accounts for the difference in length between the two proteins: SusG has a maximum length of ~ 120 Å due to the extended CBM58, while BoGH13A$_{Sus}$ is ~ 100 Å despite harboring two N-terminal domains.

The GH13 domains (509 amino acids in SusG and 448 in BoGH13A$_{Sus}$) align with a rmsd of 1.9 Å, which aligned 71% of the structure (Fig. 3C). Alignment of the amylase folds of these proteins revealed that the orientation of the proteins in the outer membrane is inverted (Fig. 3D). Both proteins have SPII signals [47]; SusG is lipidated at an N-terminal cysteine to anchor it on the cell surface and mutation of this cysteine to alanine abrogates surface exposure [48]. To confirm that BoGH13A$_{Sus}$ is lipidated in a similar manner, the equivalent cysteine was mutated to alanine. Cells harboring this allele produce BoGH13A$_{Sus}$ but fail to grow on starch, suggesting the protein is not translocated to the outside of the cell (Supplementary Fig. 4). If each protein has their N-terminal Cys proximal to the membrane, we see that BoGH13A$_{Sus}$ and SusG have opposite orientations making BoGH13A$_{Sus}$ "flipped" in the outer membrane compared to SusG (Fig. 3D). Thus, BoGH13A$_{Sus}$ is oriented such that the CBMs are proximal to the membrane and the amylase C domain is distal to the membrane, whereas SusG is oriented with the GH13 C domain proximal and the CBM58 distal to the cell surface. However, for SusG there are 40 residues between the lipidated cysteine and the start of the A domain; thus, there may be some flexibility in how the catalytic domain is positioned relative to the outer membrane. Similarly, there are 30 residues between the lipidated cysteine and the start of CBM98 in BoGH13A$_{Sus}$.

The most significant structural difference between the catalytic domains of BoGH13A$_{Sus}$ and SusG lies within the B domain. The B domain in BoGH13A$_{Sus}$ is much smaller, made up of ~ 47 amino acids compared to ~ 89 in SusG. The bulkiest part of the SusG B domain deviates from the active site toward the CBM58 (Supplementary Fig. 5). While the B domains shape the active sites in both proteins, in BoGH13A$_{Sus}$, it also contacts the CBMs. Overall, it appears that the bulkier secondary structures missing from the BoGH13A$_{Sus}$ B domain are made up for by the presence of CBM98 and CBM48. Both B domains donate an aspartate and asparagine to coordinate an ion next to the active site.

BoGH13A$_{Sus}$ coordinates Mn^{2+} and Ca^{2+}

Amylases typically coordinate at least one Ca^{2+} ion to maintain structural integrity of the protein [49]. Both BoGH13A$_{Sus}$ and SusG bind one Ca^{2+} ion on the surface of the A domain, in different positions, and for BoGH13A$_{Sus}$, this Ca^{2+} is coordinated by E274 of CBM48. Many amylases also coordinate a metal ion between the A and B domains near the active site. SusG binds Ca^{2+} at this site via side-chain carboxylates and backbone carbonyl oxygens, while the BoGH13A$_{Sus}$ utilizes the His437 imidazole N as part of the coordination sphere, precluding Ca^{2+} as the likely metal. (Supplementary Fig. 6A) [50]. The metal is octahedrally coordinated, and although Mg^{2+}–His interactions are rare, they are not unprecedented [51]. Attempts to remove this metal using EDTA or 1,10-phenanthroline did not influence activity. Therefore, we added Mg^{2+}, Mn^{2+}, Co^{2+}, Cu^{2+}, Ni^{2+}, or Zn^{2+} to untreated BoGH13A$_{Sus}$ to assess their effects on starch hydrolysis. Mg^{2+} and Mn^{2+} were the only metals that did not impair activity at the highest concentrations tested (Supplementary Fig. 6B).

To identify the bound ion, we employed native mass spectrometry (MS) and inductively coupled plasma MS (ICP-MS), and both strongly suggest that the metal near the BoGH13A$_{Sus}$ active site is Mn^{2+} (Fig. 3E and Supplementary Fig. 6C,D). The leftover native mass once the bound waters and Ca^{2+} were subtracted suggested that a single Mn^{2+} ion or Ca^{2+} and water were present (Supplementary Fig. 6E). However, ICP-MS confirmed the presence of both Mn^{2+} and Ca^{2+} bound to BoGH13A$_{Sus}$ (Fig. 3E). Therefore, based on these data and the known average coordination

geometry for various protein-bound metals [52], we propose that BoGH13A$_{Sus}$ utilizes a Mn^{2+} at this conserved ion-binding site. Mg^{2+} consistently refined with residual F_o–F_c density, whereas Mn^{2+} did not (Supplementary Fig. 6F, G). Both the Mn^{2+} and the Ca^{2+} afforded metal–ligand distances in line with structures in the Cambridge Structural Database according to the CheckMyMetal server [53] (Supplementary Fig. 6H, I). To our knowledge, the use of Mn^{2+} is unprecedented for an α-amylase as others are known to use no metal, Ca^{2+}, a dual Ca^{2+}, Na^{2+}, a Ca^{2+}/Na^{2+}/Ca^{2+} triad or a dual Ca^{2+}/Zn^{2+} strategy in the case of those from hyperthermophilic organisms [54–58].

BoGH13A$_{Sus}$ E523Q maltoheptaose-bound structure reveals two additional binding sites

To evaluate maltooligosaccharide binding to BoGH13A$_{Sus}$, we created site-directed mutant E523Q, the putative catalytic proton donor, which rendered the enzyme inactive for hydrolysis on a blocked pNP-G7 substrate (data not shown). Blocked pNP-G7 harbors an ethylidene moiety at the non-reducing end, making it an unsuitable α-glucosidase substrate. Once acted on by BoGH13A$_{Sus}$, however, the new non-reducing end becomes an α-glucosidase substrate, leading to the liberation of pNP. The structure of BoGH13A$_{Sus}$ E523Q with maltoheptaose (G7; 2.1 Å resolution, R_{work}/R_{free} = 17.7%/22.3%, Table 1) and the native enzyme structure overlay with an rmsd of 0.3 Å, suggesting limited flexibility in the enzyme. In addition to the active site, electron density for maltooligosaccharide was observed at both CBM98 and a surface binding site (surf site) (Fig. 4A).

CBM98 (residues 44–163) adopts a classic β-sandwich fold with the same pattern of connectivity (β1, β2, β4, β5 on one face and β3, β6, β7 on the opposite face) as starch-binding CBM48 and CBM21 despite having different binding sites (Fig. 4B). There was sufficient electron density to model in five glucose residues at the CBM98 binding site (Fig. 4C). The CBM98 binding platform is formed by two aromatic residues, W92 and W98, along with several polar residues and carbonyl backbones that primarily hydrogen bond with O2 and O3 of the glucopyranose rings. K275 from the neighboring CBM48 hydrogen bonds with the O6 and O1 atoms at the non-reducing end glucose (Fig. 4D). The $2F_o$–F_c density for K275 in all four monomers in the asymmetric unit is unambiguous perhaps because it is directly adjacent to E274 which coordinates the Ca^{2+} ion. It is possible that this Ca^{2+} provides structural integrity for the recognition of longer maltooligosaccharides. Overall, while CBM98 displays some features common to all CBMs including starch-binding CBMs, it has been classified as a new family based on lack of sequence similarity and correspondingly low structural similarity. Unlike CBM98, we did not observe bound maltooligosaccharide to the CBM48 in our structures.

The BoGH13A$_{Sus}$ E523Q structure with G7 enabled identification of a surface binding site on the catalytic domain positioned between α6 and β7 (W555) and a short helix between β7 and α7 (Y592) (Fig. 4E, F and Supplementary Fig. 7A). Surface starch binding sites are found within a number of GH13s and like CBMs are believed to enhance enzyme efficiency by concentrating substrate near the active site [61]. Sufficient electron density allowed for fitting five glucopyranoses including one with partial density at the reducing end (Fig. 4E). Three glucoses make direct hydrogen bonds to polar residues with indirect interactions mediated by water molecules. SusG also has a surface starch-binding site, but it is found on α6a and β6b and is ~ 12 Å from the BoGH13A$_{Sus}$ surf site (Supplementary Fig. 7B, C). In both enzymes, the reducing end of the bound oligosaccharide at the surface site is directed toward the reducing end of oligosaccharide exiting the active site. For BoGH13A$_{Sus}$, the distance between the reducing ends of the modeled maltooligosaccharides at these two sites is 12.7 Å, while in SusG the distance between the O1 reducing ends is 8.5 Å (Supplementary Fig. 7D).

The active site of BoGH13A$_{Sus}$ accommodates α1,6 linkages

Within the active site of the E523Q mutant, we observed a transglycosylation product in all four monomers resulting in an α1,6 bond. Since these crystals grew at pH 8.5, it is possible that water served as a proton donor in place of Glu. Other possibilities include a low level of deamidation producing a small fraction of wild-type enzyme or translational misincorporation (GAA→CAA), as has been observed in other glycoside hydrolases [62]. However, the E523Q mutant showed no activity on the blocked pNP-G7 substrate after 20 min (data not shown). To further probe this, reactions mimicking crystallization conditions were set up wherein 10 mg/ml (120 µM) of WT and E523Q BoGH13A$_{Sus}$ were incubated overnight at room temperature in crystallization buffer with 10 mM G7. Equivalent reactions containing 1 µM of WT and E523Q BoGH13A$_{Sus}$ were set up for comparison. 1 µM WT produced transglycosylation products in the crystallization buffer but not at pH 7 at 37 ºC, and 120 µM E523Q showed a small amount of activity (Supplementary Fig. 8A). Notably, the blocked pNP-G7 kinetics were performed with only 150 nM enzyme. Maltogenic amylases are prone to transglycosylation, specifically α1,6-transglycosylation because newly formed α1,4 bonds are readily hydrolyzed but α1,6 bonds are not [63, 64]. Taken together, these data suggest that either a small amount of WT enzyme in the 120 µM batch of E523Q or low-level activity of the mutant enzyme during the two

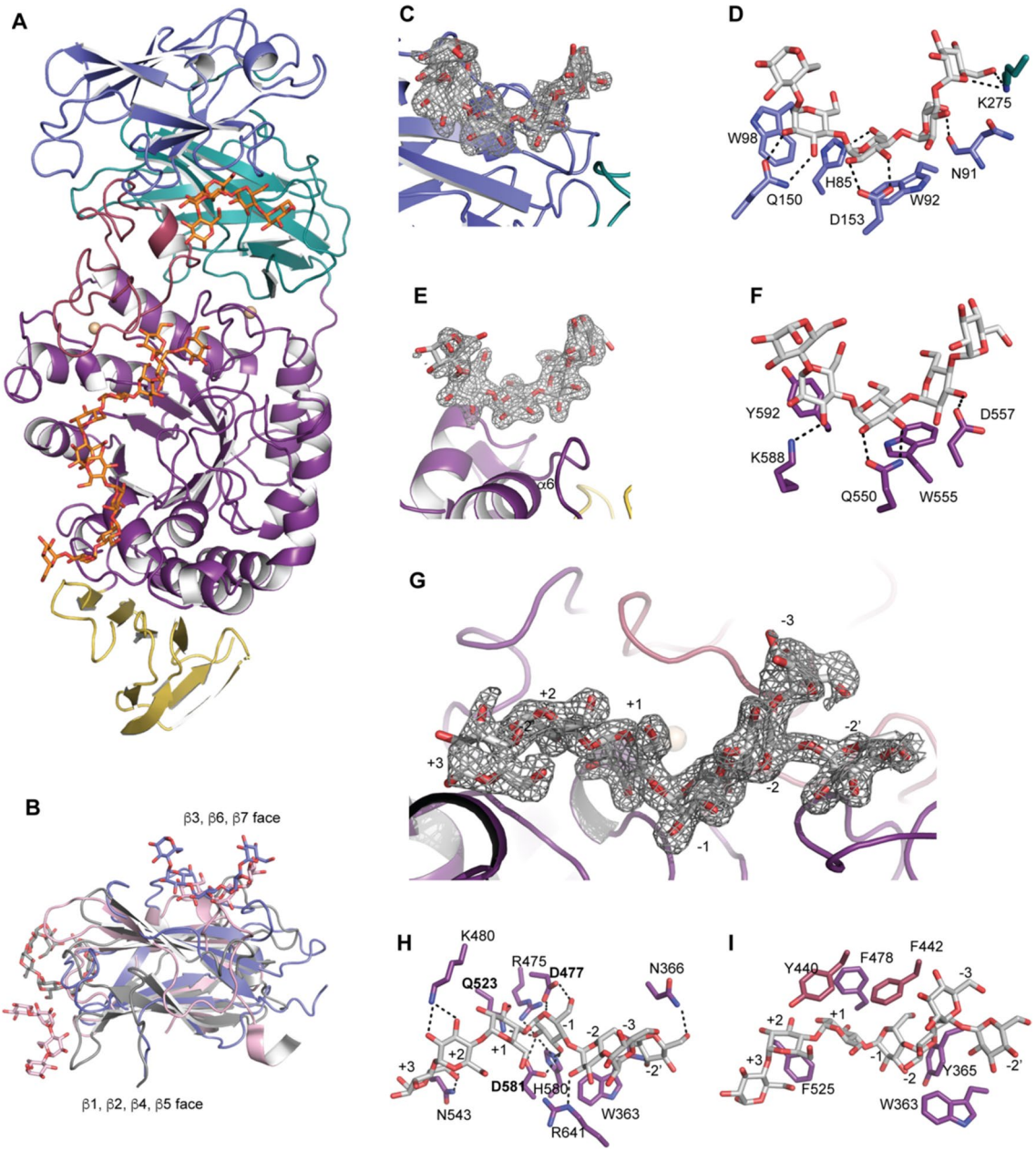
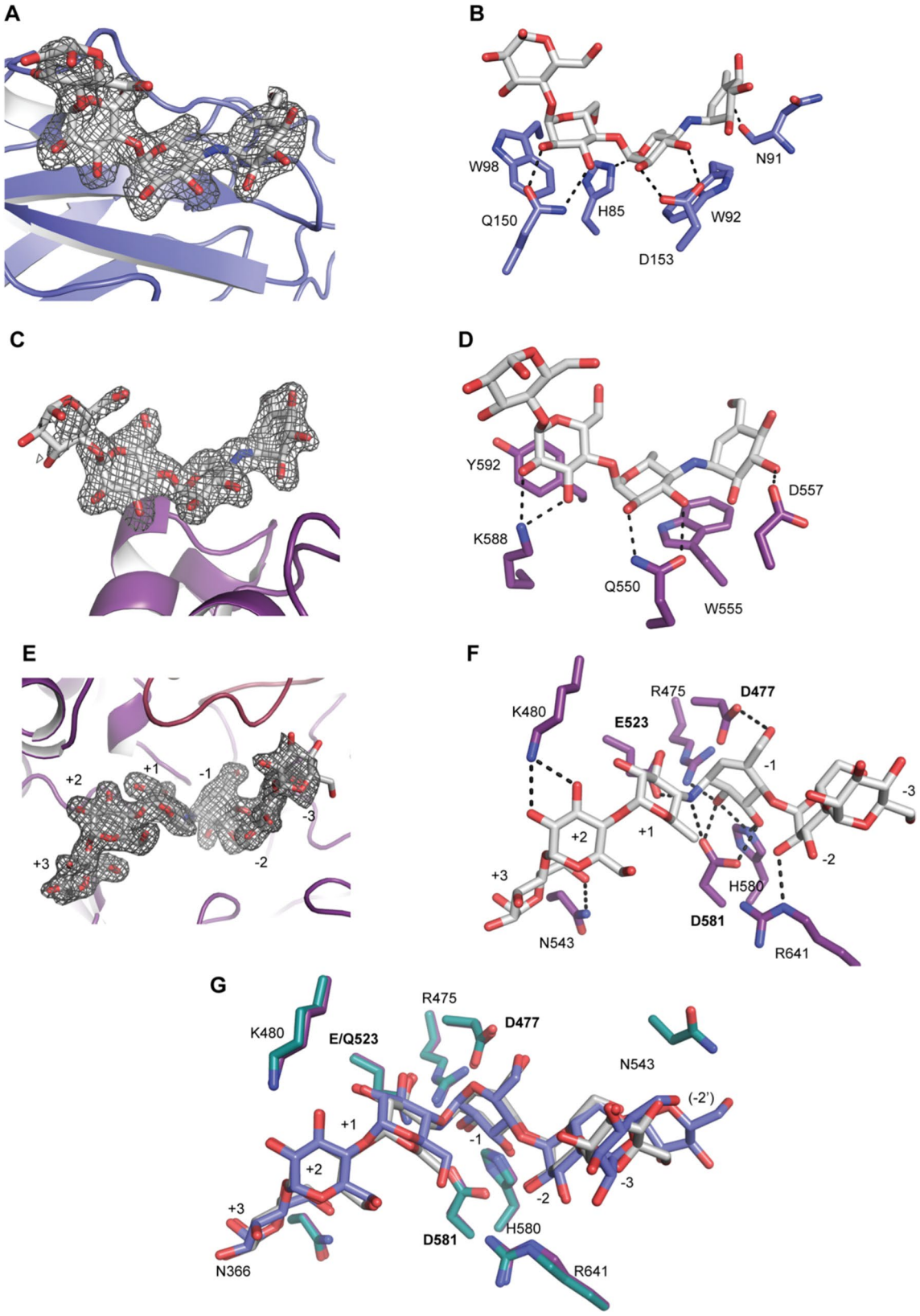


Fig. 4 Maltoheptaose bound to BoGH13A$_{Sus}$ E523Q. **A** Model of full-length BoGH13A$_{Sus}$ bound to maltoheptaose (G7). Domains are colored as in 3A with glucoses shown as orange sticks. Monomer A is displayed. **B** An alignment of BoGH13A$_{Sus}$ CBM98 from monomer D (slate), CBM48 from *Cyanothece* sp. ATCC 51142 branching enzyme (PDB ID: 5GQV, pink [59]) and CBM21 from *Rhizopus oryzae* glucoamylase (PDB ID: 2V8M, gray, [60]. Bound oligosaccharides are shown in the same color, as sticks, as the respective protein. **C** F_o–F_c density for oligosaccharide bound to CBM98 in monomer D. The density is contoured at 3σ. **D** The aromatic-binding platform in CBM98, comprised of W92 and W98, along with polar residues involved in hydrogen bonding (within 3.2 Å) and the hydrogen bonds themselves are shown. **E** F_o–F_c density for oligosaccharide bound to the surface binding site (surf site) of monomer A. The density is contoured at 3σ. **F** The aromatic-binding platform at the surf site, polar residues involved in hydrogen bonding (within 3.2 Å) and the hydrogen bonds are shown. **G** F_o–F_c density for oligosaccharide bound in the active site of monomer A. Subsites from the reducing (+ 3) to the non-reducing end (− 3) are labeled. The − 2 branch point Glc is − 2′. **H** Hydrogen-bonding (within 3.2 Å) network in the active site. Amino acids labeled in bold constitute the catalytic triad. **I** Aromatic residues contributing to the active site architecture are displayed. Those colored in purple are from the A domain and those in burgundy are from the B domain

weeks when crystals grew may have resulted in transglycosylation. For example, the small oligosaccharides produced overnight by 120 µM E523Q may have eventually served as transglycosylation substrates (Supplementary Fig. 8A). It is worth noting that an active site mutant of SusG crystallized with glucosyl–α1,6-maltotriosyl–α1,6-maltotriose yielded


A
B
W98
N91
Q150
H85
W92
D153
C
D
Y592
D557
K588
Q550
W555
E
+2
+1
-1
-3
+3
-2
F
K480
R475
D477
E523
-1
-3
+2
+1
+3
-2
H580
D581
N543
R641
G
R475
K480
E/Q523
D477
N543
(-2')
+1
-1
+2
-3
+3
-2
H580
D581
N366
R641

◂**Fig. 5** Acarbose bound to BoGH13A$_{Sus}$. **A** F_o–F_c density for acarbose bound to CBM98 in monomer C. The density is contoured at 3σ. **B** The aromatic-binding platform in CBM98, comprised of W92 and W98, along with polar residues involved in hydrogen bonding (within 3.2 Å) and the hydrogen bonds are shown. **C** F_o–F_c density (contoured at 3σ) for acarbose bound to the surface binding site (surf site) of monomer A. **D** The aromatic binding platform at the surf site, along with polar residues involved in hydrogen bonding (within 3.2 Å), and the hydrogen bonds are shown. **E** F_o–F_c density for the acarbose-derived hexasaccharide bound in the active site of monomer A. Subsites are labeled according to Fig. 4G. **F** Hydrogen bonding (within 3.2 Å) network in the active site. Amino acids labeled in bold constitute the catalytic triad. **G** Overlay of G7-bound and acarbose-bound BoGH13A$_{Sus}$ structures. The acarbose derived hexasaccharide is shown as white sticks and G7 derived oligosaccharide shown in slate. Residues from the acarbose bound structure are colored in purple, whereas those from the G7 bound structure are colored in teal

electron density for a transglycosylated product due to trace amounts of WT enzyme [35].

F_o–F_c electron density was sufficient to model a partial molecule of G7 from subsites + 3 to − 3 plus one glucose residue appended to the − 2 subsite glucose via an α1,6 bond in the active site of chain A (Fig. 4G). In chain B, there was sufficient density at the branch point to model 2 Glcs (Supplementary Fig. 8B). In chain B, the non-reducing end at the − 2″ position is just 14.0 Å from the O1 reducing end of the oligosaccharide bound to CBM98, while the O4 of Glc at the − 3 subsite is 15 Å from the O1 reducing end (Supplementary Fig. 8C). This raises the intriguing possibility that a longer polysaccharide could potentially bind and span the two sites. The catalytic triad was readily identified as D477, Q(E)523 and D581 along with additional conserved residues in GH13s including R475 and H580 [65] (Fig. 4H).

While there are no hydrogen bonds to the − 3 glucose since it extends into solvent, the − 2 and − 2′ Glcs are held in place by hydrogen bonding to R641 as well as the amide backbone of W363 and side chain of N366, respectively (Fig. 4H, Supplementary Fig. 8D). In chain B, the backbone carbonyl and side chain of T642 also hydrogen bond with the − 2″ glucose at position O2. (Supplementary Fig. 8D). Although W363 does not pi stack with the − 2′ glucose, it does constrict its positioning (Fig. 4I and Supplementary Fig. 8D.) Overall, the − 2′ Glc resulting from the α1,6 linkage is recognized by BoGH13A$_{Sus}$ more than the − 3 Glc, the latter of which makes no hydrogen-bonding or pi stacking interactions with the enzyme. This is somewhat analogous to what is observed in the SusG structure with GMM as an α1,6 bond is found in the same position as in BoGH13A$_{Sus}$, at the − 2 subsite. In SusG, there is a dedicated tryptophan, W349, whose side chain accommodates the − 2′ glucose [35].

Several aromatic residues (Y365, F478, Y440, F442 and F525) facilitate active site positioning of substrate at the − 1, + 1 and + 2 subsites (Fig. 4I). The BoGH13A$_{Sus}$ + 1 and + 2 binding determinants Y440 and F525 are akin to those described for maltogenic amylases of the subfamily GH13_36 [40] (Supplementary Fig. 8E). This supports the TLC activity data in that maltose is the smallest maltooligosaccharide released. While SusG is part of this subfamily, it is an outlier in that its + 1 and + 2 binding determinants are much smaller, a histidine and leucine, explaining why only the + 1 subsite has to be occupied for catalysis to occur (Supplementary Fig. 8E) [23, 40]. The + 2 and + 3 glucoses in BoGH13A$_{Sus}$ are anchored by K480 and N543, respectively (Fig. 4H). Additionally, the BoGH13A$_{Sus}$ B domain interacts much more with the substrate reducing end than the non-reducing end, while the opposite is true of the SusG domain (Supplementary Fig. 8F). Overall, the wide active site of BoGH13A$_{Sus}$ can accommodate long α-glucan chains, as seen in activity assays, though it seems that the reducing end (positive subsites) contributes more toward anchoring the substrate in the hydrolytic or transglycosylation reaction.

Structure of acarbose-bound BoGH13A$_{Sus}$

Acarbose is an FDA-approved medication used to treat type 2 diabetes. Its primary function is to inhibit extracellular glucosidases and amylases in the upper digestive tract by functioning as a transition state mimic of α1,4-glycosidic bond hydrolysis [66–68]. Acarbose is used frequently in structural studies of eukaryotic and bacterial α-amylases alike [23, 69–72]. An acarbose-bound structure of wild-type BoGH13A$_{Sus}$ was solved to a resolution of 1.99 Å (R_{work}/R_{free} of 17.5%/21.9%, Table 1). There were no large structural shifts upon binding, as each of the four monomers overlays with the G7-bound structure (rmsd of ~ 0.3 Å).

Acarbose was modeled at CBM98 and some of the surface binding sites within the four chains in the asymmetric unit (Fig. 5A–D). The binding determinants in both sites were identical as in the G7-bound structure, except that the ligand was shortened at the non-reducing end since acarbose is a pseudotetrasaccharide and thus a smaller molecule. In some subunits, there was not sufficient density to model a full acarbose at the surface site but there was partial density in chain A (Fig. 5C). This may be due to hydrolysis during crystallization, as evidenced by the presence of a trisaccharide upon resolution of crystallization drops via thin layer chromatography (TLC) (Supplementary Fig. 9).

Acarbose underwent transglycosylation during crystallization, and this product was only observed in the active site. Unlike during G7 crystallization, an α1,4 bond was introduced. Identifying transglycosylation products was complicated by the fact that these crystals grew out of a well solution with a mix of monosaccharides including glucose. TLC of drops containing crystals revealed oligosaccharides larger than acarbose (Supplementary Fig. 9A). In chain B, the density for additional pyranoses corresponded best with two Glcs added at the acarviosine reducing end. Thus, one ligand spanning subsites

−3 to + 3, with the non-hydrolyzable amino-bridge from acarviosine spanning the + 1/− 1 site was modeled (Fig. 5E, F). We cannot be certain if glucose from the crystallization solution or maltose from acarbose breakdown, performed by some maltogenic amylases, was utilized in the transglycosylation reaction. Regardless, the product is positioned similarly in the active site as G7 from subsites − 2 to + 3, but the orientation of the -3 glucose is different, likely since this residue extends into the solvent (Fig. 5G). Small changes in the hydrogen bonding network between these ligands occur due to the amino-bridge and loss of O5 at the acarbose non-reducing end, though this product fits well within the active site (Fig. 5F).

CBM98 and the surface binding site differentially influence enzymatic activity

To determine how the CBM98 and CBM48 contribute to activity, we first measured substrate binding to CBM98 via isothermal titration calorimetry (ITC) with soluble substrates. A summary of tested constructs is provided in Fig. 1D. As predicted from our crystal structure, CBM98 harbors one binding site that accepts G7, αCD and βCD equally well with K_ds of ~ 25–40 μM (Table 2, Supplementary Fig. 10). CBM98 does not bind as well to G3, which is often the minimum length for many starch targeting CBMs, and binds potato AP with similar affinity to G7. To eventually test the role of CBM98 in hydrolysis, we mutated W92 and W98 to Ala and this eliminated binding to any oligosaccharide or polysaccharide (Table 2, Supplementary Fig. 11). The presence of CBM48 does not affect binding affinity to or the number of binding sites for maltooligosaccharides as a CBM98–CBM48 construct bound similarly to a CBM98 construct alone. Conversely, CBM98–CBM48 demonstrates a tenfold decrease in binding affinity to amylopectin compared to CBM98 alone (Table 2, Supplementary Fig. 12).

We were unable to isolate CBM48 on its own, so a CBM98–CBM48 construct with the W92/98A mutations was tested and lacked binding to maltooligosaccharides (Table 2). We note that very modest isotherms were present in αCD and βCD titrations, but these could not be modeled (Supplementary Fig. 13). This validated our previous structural observations that CBM48 is a non-binding CBM. That CBM98–CBM48 bound poorly to amylopectin suggests that the CBM48 may somehow restrict the CBM98 binding site or impose steric restraints on CBM98 itself.

To ascertain the roles of CBM98 and the surface binding site in polysaccharide binding and catalysis, native PAGE and end point kinetics using wild-type and variant enzymes were performed. Three variant BoGH13A$_{Sus}$ proteins were constructed: CBM* (W92, 98A), surf* (W555, Y592A) and CBM*surf* (W92/98/555A and Y592A) (Fig. 1D). While we could not confirm the role of the surface binding site amino acids via ITC due to binding at CBM98 and the active site, similar mutagenesis of aromatics in SusG led to substantially different enzyme turnover [23]. To determine overall starch binding by the BoGH13A$_{Sus}$ mutants, we performed affinity PAGE with 1 mg/ml (0.1%) potato AP, glycogen, pullulan or autoclaved potato starch (PS) (Fig. 6A). The smiling observed in the wild-type band on all four polysaccharides is due to enzymatic breakdown of carbohydrate while running the gel [61, 74]. Overall, the migration pattern of each enzyme variant was different, supporting that both the CBM98 and surface site contribute to substrate recognition. WT BoGH13A$_{Sus}$ displays the lowest affinity for glycogen and similar affinity for AP, pullulan and PS (Fig. 6A). Interestingly, the CBM*surf* mutant displays loss of binding on all substrates tested with the individual mutants showing intermediate binding based upon the type of starch. During PS, AP and pullulan binding, the CBM* mutation appears to drive the loss of binding in the CBM*surf* double mutant. Glycogen binding is affected most by the surf* mutation but is partially restored by the introduction of the CBM* variation (Fig. 6A). The presence of multiple bands in some of the carbohydrate containing gels suggests that BoGH13A$_{Sus}$ may have slightly different affinities for different polysaccharide features (i.e., branch points, reducing ends, non-reducing ends, amylopectin double helices). The

Table 2 BoGH13A$_{Sus}$ CBM98 and CBM98–CBM48 ITC data

Construct	Ligand	K_d (μM) ± SD	N ± SD
CBM98	maltoheptaose	32 ± 9.8	0.8 ± 0.16
	α-cyclodextrin	42 ± 13	0.75 ± 0.2
	β-cyclodextrin	26 ± 3.8	0.81 ± 0.03
	maltotriose	346 ± 46	1*
	AP potato	71 ± 3	1.01 ± 0.01#
CBM98 W92,98A	maltoheptaose	NB	NB
	α-cyclodextrin	NB	NB
	β-cyclodextrin	NB	NB
	maltotriose	NB	NB
	AP potato	NB	NB
CBM98–CBM48	maltoheptaose	33 ± 9.9	0.74 ± 0.26
	α-cyclodextrin	43 ± 18	0.58 ± 0.3
	β-cyclodextrin	21 ± 1.4	0.77 ± 0.09
	maltotriose	360 ± 102	1*
	AP potato	976 ± 23	0.99 ± 0.03#
CBM98–CBM48 W92,98A	maltoheptaose	NB	NB
	α-cyclodextrin	NB	NB
	β-cyclodextrin	NB	NB
	maltotriose	NB	NB
	AP potato	NB	NB

NB stands for no binding

*Indicates that N was set to 1 because of low-affinity binding

#Indicates that substrate concentration was adjusted to achieve an *N* of 1 according to [73]

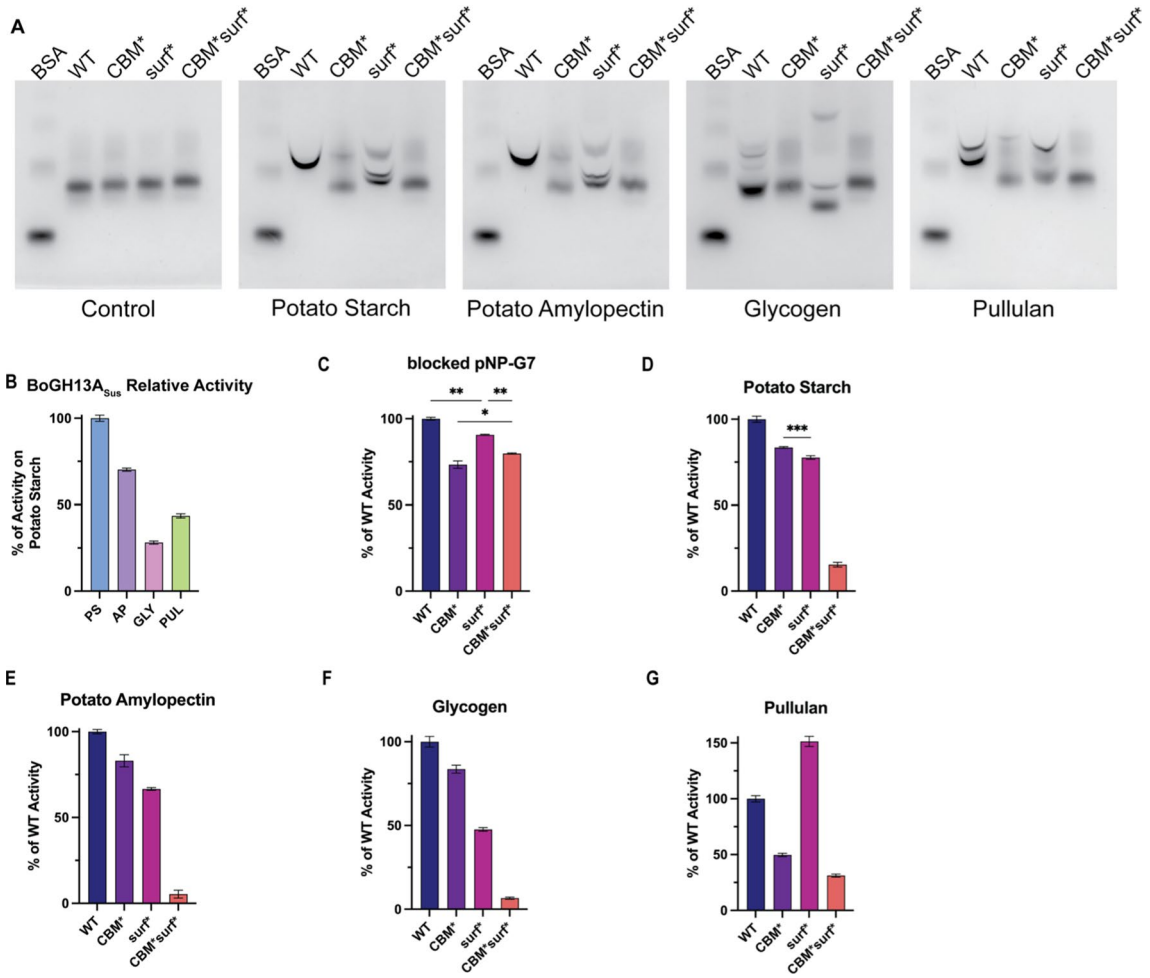


Fig. 6 BoGH13A$_{Sus}$ activity on polysaccharides. **A** WT and variant BoGH13A$_{Sus}$ were subjected to native gel electrophoresis in the absence (control gel) or presence of 0.1% of the indicated polysaccharides in a 6% gel. Bovine serum albumin (BSA) was loaded as a negative control. Gels were visualized with Coomassie. **B** WT BoGH13A$_{Sus}$ relative activity on polysaccharides was assessed using the DNSA method. PS = autoclaved potato starch; AP = potato amylopectin, GLY = glycogen, PUL = pullulan. 2.5 mg/ml carbohydrate and 100 nM BoGH13A$_{Sus}$ were used at room temperature for a 10 min reaction. BoGH13A$_{Sus}$ was most active on PS and set at 100% to compare other polysaccharides. Statistical analysis was performed using an ordinary one-way ANOVA comparison. All comparisons were statistically significant with an adjusted $p < 0.0001$ (****). **C** WT and variant BoGH13A$_{Sus}$ relative activity on an oligosaccharide was assessed at room temperature using an amylase kit from LSBio. 150 nM enzyme was used. WT was most active and set at 100% for comparison to other variants. Comparisons for which statistics are not shown were significant with an adjusted $p < 0.0005$ (***). **$p < 0.005$. *$p = 0.019$. **D–G** The DNSA method was used to characterize relative activity of variant BoGH13A$_{Sus}$s to WT BoGH13A$_{Sus}$ (set to 100% activity on each of the indicated polysaccharides). 100 nM enzyme at room temperature was used for all reactions. All comparisons were statistically significant with an adjusted $p < 0.0001$ (****) except for the CBM*/surf* comparison on PS which had an adjusted $p = 0.0001$ (***)

multi-band patterning with the variants could be due to an overall lower affinity for polysaccharide that leads to a bound and unbound fraction. All four enzymes tested run similarly on the control gel, suggesting the multiple bands are most likely not due to misfolded protein.

WT BoGH13A$_{Sus}$ activity on 2.5 mg/ml (0.25%) of the same carbohydrates tested above nearly follows the binding preferences wherein the enzyme is most active on autoclaved PS, followed by AP, pullulan, and is least active on glycogen (Fig. 6B). With an oligosaccharide substrate, blocked pNP-G7, the CBM* mutant is 75% as active as the WT enzyme. This suggests that CBM98 may assist in loading and/or positioning of substrate in the active site. The surf* variant is nearly as active as WT on blocked pNP-G7, while the CBM*surf* protein displays intermediate activity compared to the single mutations (Fig. 6C).

On all four polysaccharides, the CBM*surf* double mutant is nearly inactive, while the single mutants retain some activity depending on the substrate. Despite the CBM98 contributing most to binding PS and AP, the CBM*

mutant retains more activity on these substrates than the surf* mutant (Fig. 6D, E). The surf* variant is less active on glycogen than the CBM* variant, which correlates with its binding activity (Fig. 6F). Finally, the surf* enzyme is 1.5 × as active on pullulan as the WT enzyme, but this is completely offset in the CBM*surf* double mutant (Fig. 6G). Since the BoGH13A$_{Sus}$ surface binding site accommodates ligand such that the reducing end of bound substrate would clash with the reducing end of active site bound substrate, it may be beneficial during pullulan hydrolysis to eliminate non-productive binding at the surface binding site. It is unclear why the same effect is not observed with other polysaccharides except that pullulan is linear, so the surface binding site could be less dispensable for more complex, branched polysaccharide recognition.

Complementation of α-amylases between species is manganese dependent

To understand whether the activities of extracellular α-amylase enzymes from Bt and Bo can compensate for one other in the context of otherwise very similar PUL, Bt*ΔsusG* and Bo*Δbovatus_03803* were complemented with α-amylase from the opposite species. *bovatus_03803* encodes BoGH13A$_{Sus}$. Bt*ΔsusG* and Bo*Δbovatus_03803* grow on glucose, maltose and G7, but not on polysaccharides ([21, 22, 24], Supplementary Fig. 14A.) Bo*Δbovatus_03803* expressing *susG* within BoSus grows similarly to WT Bo on potato starch, glycogen and pullulan and better than WT on amylopectin (Fig. 7A, Supplementary Fig. 14B). The same is not true when Bt*ΔsusG* expresses *bovatus_03803* from within BtSus. Bt*ΔsusG* + *bovatus_03803* exhibits a significant growth defect on pullulan and potato starch and cannot grow on glycogen or amylopectin (Fig. 7A, Supplementary Fig. 14B). However, the "swapped" mutants display

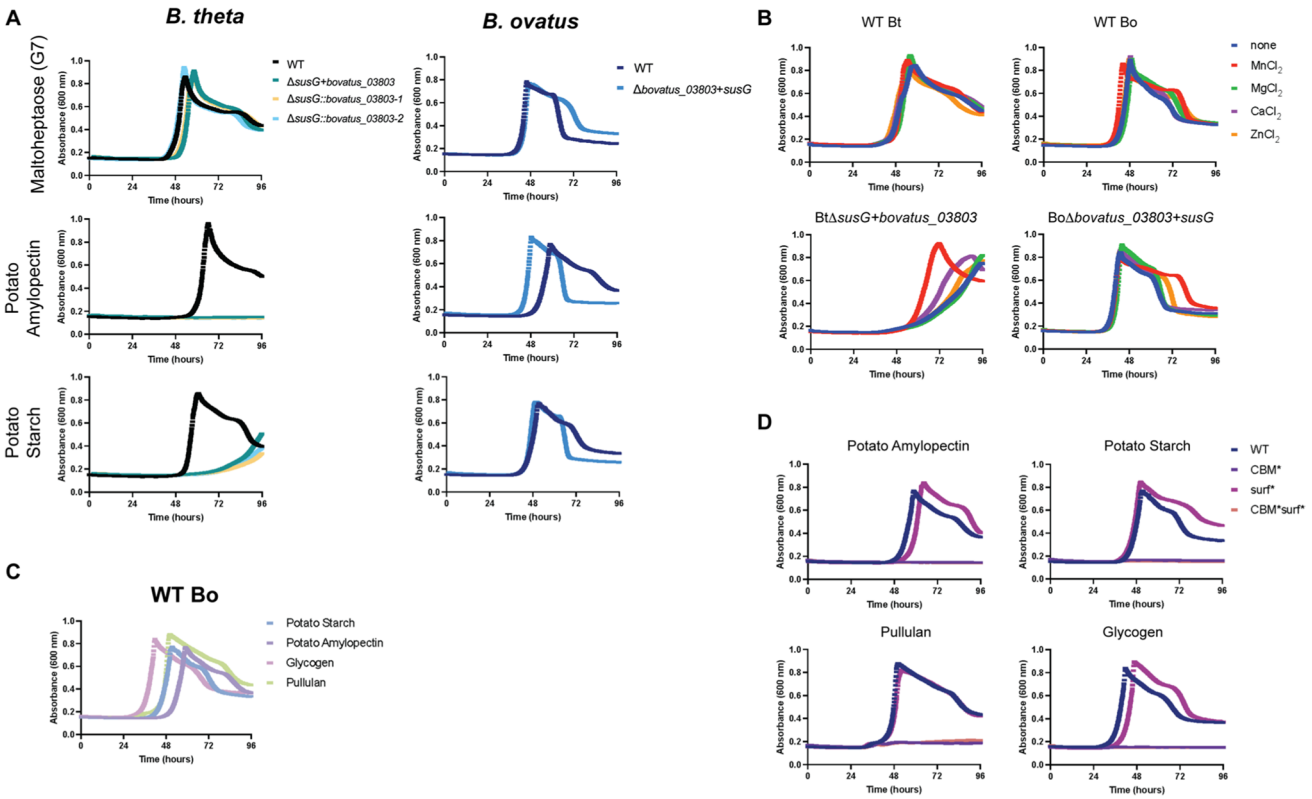


Fig. 7 Bo*Δbovatus_03803* + *susG* grows on starch, while Bt*ΔsusG* + *bovatus_03803* growth on starch is enhanced by $MnCl_2$. For all growths, bacterial strains were pre-grown on minimal media (MM) + 5 mg/ml glucose. The following day, they were washed in 2 × MM and back diluted 1:100 into MM + 2.5 mg/ml of the indicated carbohydrates and grown in triplicate. Bt*ΔsusG* + *bovatus_03803* harbors the *bovatus_03803* gene in place of the *susG* gene. The *ΔsusG::bovatus_03803-1* and *-2* strains express the *bovatus_03803* gene from one of two sites under the control of a constitutive promoter [75]. **A** WT Bt and Bo and α-amylase swap strains grown on maltoheptaose, potato amylopectin and potato starch. **B** The same strains were grown in potato starch with or without 100 μM of supplemented metal. **C** WT Bo was grown on the indicated polysaccharides. **D** WT and variant Bo strains were grown on the indicated polysaccharides

WT growth on glucose, maltose and G7 (Fig. 7A, Supplementary Fig. 14C), indicating that other Sus proteins were not affected by this genetic manipulation. In these experiments, the α-amylase encoding gene was inserted into the chromosome at its native location in either species. We also tested complementation by constitutive expression of *bovatus_03803* from two regions of the Bt chromosome, and this does not rescue the Bt*ΔsusG* + *bovatus_03803* growth lag (Fig. 7A, Supplementary Fig. 14B).

Furthermore, BoGH13A$_{Sus}$ is expressed by Bt as demonstrated by Western blot analysis of whole cells grown on maltose (Supplementary Fig. 15). Unfortunately, numerous attempts to visualize BoGH13A$_{Sus}$ on the outside of Bt (or in its native context on the outside of Bo) via immunostaining with the αBoGH13A$_{Sus}$ antibody were unsuccessful. Nonetheless, SusG is also expressed by Bo when grown on maltose (Supplementary Fig. 15). These data collectively suggest that the growth lag in Bt*ΔsusG* + *bovatus_03803* may be due to suboptimal BoGH13A$_{Sus}$ presentation at the cell surface, suboptimal association with BtSus proteins or that BoGH13A$_{Sus}$ starch processing is suboptimal for import via the BtSus machinery.

The Western blot analysis of whole cells suggests that Bt can produce BoGH13A$_{Sus}$ but does not indicate whether the protein is folded properly. Furthermore, Bt does not normally make a Mn^{2+}-dependent α-amylase, so we wondered whether $MnCl_2$ supplementation in the growth media could rescue the growth lag in Bt*ΔsusG* strains complemented with BoGH13A$_{Sus}$. To our surprise, $MnCl_2$ and to a much lesser extent $CaCl_2$ supplementation rescued Bt*ΔsusG* + *bovatus_03803* growth on potato starch and pullulan but $MgCl_2$ or $ZnCl_2$ did not (Fig. 7B, Supplementary Fig. 16A). WT Bt growth was not affected by supplemented cations, and importantly, growth on glucose and pectic galactan or rhamnogalacturonan I, which does not require SusG or an extracellular Mn^{2+}-dependent enzyme, was not enhanced (Supplementary Fig. 16B-D) [76, 77]. WT Bo growth in potato starch and pullulan was slightly enhanced with supplemented $MnCl_2$ but not by other cations, but the Bo*Δbovatus_03803* + *susG* strain did not benefit from $MnCl_2$ supplementation (Fig. 7B, Supplementary Fig. 16A). Although we do not know the metal dependence of the extracellular Bo enzymes required for growth on these substrates, growth of either Bo strain on pectic galactan and rhamnogalacturonan I was not enhanced by any cations (Supplementary Fig. 16B, C). Remarkably, both Bo strains exhibited a marked growth defect in glucose in the presence of $MnCl_2$, but not in the presence of other cations (Supplementary Fig. 16D). It is possible that Bo has a more robust Mn^{2+} import system than Bt that leads to these deleterious effects. Given that glycolytic enzymes can exhibit different metal requirements, it is possible that Bt uses primarily Mg^{2+} ones and Bo uses Mn^{2+} containing ones [78].

CBM98 binding is required for growth on polysaccharides

While BoGH13A$_{Sus}$ is least active on glycogen (Fig. 6B), Bo has the shortest apparent lag on this substrate (Fig. 7C). BoGH13A$_{Sus}$ may cleave glycogen in fewer places compared to other polysaccharides, thereby liberating longer oligosaccharides, fewer reducing ends and leading to lower apparent relative activity (Figs. 2, 6B). These longer oligosaccharides are apparently taken up readily by Bo and lead to a shorter overall lag time compared to other substrates (Fig. 7C).

The contributions of CBM98 and the surface binding site to growth were evaluated by complementing Bo*Δbovatus_03803* with BoGH13A$_{Sus}$ variants at the native location in the *sus* locus. CBM*, surf* and CBM*surf* proteins are all made by Bo although CBM* and CBM*surf* variants are not made at comparable levels to WT or surf* SusG (Supplementary Fig. 15). This may be due to stability issues as CBM* and CBM*surf* enzymes also exhibit a slight defect using blocked pNP-G7 as a substrate (Fig. 6C). Nonetheless, binding at CBM98 is absolutely required for growth on all polysaccharides tested as CBM* and CBM*surf* variants do not grow (Fig. 7D). All variants grow on glucose, maltose and G7 normally (Supplementary Fig. 17). The surf* strain exhibits a small growth lag on potato amylopectin and glycogen but grows like WT Bo on pullulan and potato starch. CBM*surf* is almost completely inactive on polysaccharides in vitro in the conditions tested (Fig. 6D–G), and this correlates with its inability to support growth in vivo. The CBM* variant lost 20–50% activity in vitro compared to WT BoGH13A$_{Sus}$ and yet did not restore growth.

Evolutionary diversity of CBM98 and GH13_47

We analyzed the distribution of the newly defined CBM98 and GH13_47 modules based on their annotation in Bacteroidetes species in PULDB [79] and beyond in CAZy genomes [29]. These modules are exclusively restricted to the Bacteroidetes phylum. But, they are found in all its main taxonomic classes (Bacteroidia, Flavobacteria, Cytophagia, Sphingobacteria and Chitinophaga), thus covering a wide diversity of environments besides the human gut including soil and aquatic habitats.

The closest homologs of GH13_47 form several groups of GH13 enzymes which have not been classified in any GH13 subfamily, nor been characterized to date (designated as related groups GH13_group_re1-re3 hereafter). The closest characterized relative to GH13_47 is the GH13_10 subfamily (malto-oligosyltrehalose trehalohydrolases) [80]. On the other hand, SusG belongs to GH13_36, which is related to subfamilies 23 and 31. A phylogenetic tree was constructed based only on the GH13

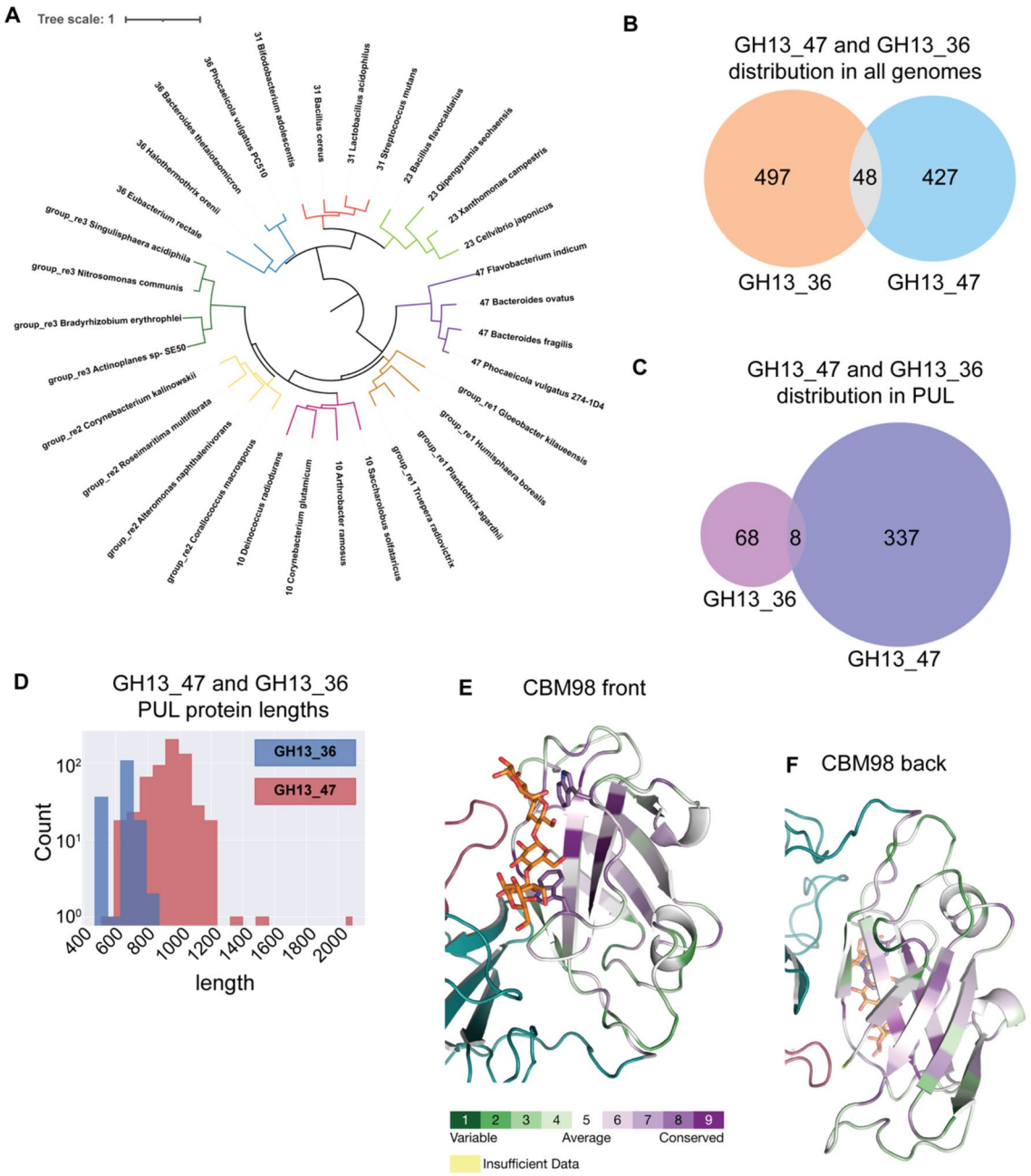


Fig. 8 Distinct evolutionary features of GH13_47, GH13_36 and CBM98. **A** A phylogenetic tree was generated from an alignment with a subset of GH13 sequences from subfamilies 23, 31, 36, 10, and group re1-3 rooted on the node separating the 10/47 and 23/31/36 branches (see Supplementary Table 3 for sequences). Branch IDs include the subfamily and organism name. **B** Instances of all genomes encoding GH13_47 and/or GH13_36 in the CAZy database. **C** Instances of subfamily GH13_47 and GH13_36 in PUL from all Bacteroidetes. **D** Histogram of the length of subfamily GH13_47 and GH13_36 proteins from all PUL. **E**, **F** An alignment of 288 CBM98 modules was submitted to the Consurf server [81-83] and mapped onto the BoGH13A$_{Sus}$-G7 structure. Residues are colored from green to purple for least to most conserved. The CBM48 is colored in teal and B domain colored in red

modules using four representatives of each group/subfamily to clarify their distinct evolutionary routes (Fig. 8A). Many similarities between GH13_47, GH13_10 and related groups can be visualized at the amino acid level by focusing on the seven conserved regions of GH13 enzymes (Supplementary Fig. 18A, Supplementary Table 3) [65].

We next estimated the prevalence of GH13_36 and GH13_47 proteins in genomes and PUL. It appears that

GH13_47 is slightly less prevalent than GH13_36 when considering all organisms, while it far outnumbers GH13_36 in Bacteroidetes genomes (Fig. 8B, Supplementary Fig. 18B). Indeed, GH13_36 is notably present in Firmicutes and several Proteobacteria phyla, raising the question as to its putative acquisition/transmission from or to other clades, in other words, its ancestral origin.

On the other hand, the GH13_47 dominance in Bacteroidetes raises the question of its prevalence in PUL and its synergy with other enzymes in these PUL. In most cases, GH13_36 and GH13_47 are predicted to belong to PUL or CAZyme clusters (59% and 75%, respectively; compare Fig. 8C with Supplementary Fig. 18B). We observed similar numbers of GH13_47 and GH13_36 in "canonical Sus" that harbor both SusA and SusB homologs (Supplementary Fig. 18C). Of note, distinct *Phocaeicola vulgatus* strains encode either GH13_47 or GH13_36 in an otherwise canonical Sus [79]. However, GH13_47 can be found in a large diversity of PUL (Supplementary Fig. 18D). We notably observe many GH13_47-containing putative starch PUL encoding a GH65 enzyme (related to maltose phosphorylases) sometimes with an additional SusA homolog. More frequent are PUL with GH13_47 as the sole CAZyme. Such PUL could still cooperate with another PUL to confer growth on starch, as many Flavobacterium species display an adjacent CAZyme cluster including both GH65 and SusA homologs. Finally, several PUL include both a GH13_47 and a GH13_36 devoid of a CBM58, as discussed hereafter (Supplementary Fig. 18D) [79].

GH13_47 proteins tend to be quite a bit longer than GH13_36 proteins in all organisms, as well as at the PUL level (Fig. 8D, Supplementary Fig. 18E). Some GH13_47 proteins, like that from *Bacteroides fragilis* (GenBank: CAH08838.1), are nearly 1000 amino acids long and contain two *Bacteroides*-associated carbohydrate-binding often N-terminal (BACON) domains at the N terminus that precede the CBM98–CBM48 [84, 85]. Also contributing to their increased length, other GH13_47 proteins harbor a type 9 secretion signal (T9SS) C-terminal domain, a system only found in the Bacteroidetes phylum [86]. The T9SS is rarely fused to a GH13_36 so may be linked to ecological niche since these signals are typically used for gliding motility or pathogenicity [86]. Several GH13_36 do not possess a CBM58, explaining the bimodal distribution of GH13_36 protein lengths (Fig. 8D).

Due to their increased length, we analyzed the domain composition of GH13_47-encoding proteins in more detail. GH13_47 modules are always found appended directly after a CBM48 (except in rare N-terminally truncated gene models). In 90% of cases, they are also preceded by a CBM98. CBM98 is almost always restricted to CBM48–GH13_47 proteins except in rare cases where it precedes a unique GH31, as exemplified by GenBank: ALO16387.1 from *Salinivirga cyanobacteriivorans*. Notably, only ~ 21% of CBM98-containing proteins mimic BoGH13A$_{Sus}$'s "simple" domain composition, while most harbor additional domains. For example, as previously mentioned, two N-terminal BACON domains were identified in GH13_47 proteins from *B. fragilis* strains, while those from *P. vulgatus* and *Bacteroides uniformis* have a single BACON domain. More importantly, ~ 62% of GH13_47 proteins have an additional conserved domain in-between the CBM98 and CBM48, as demonstrated in GenBank: AXG74135.1 [29]. This highlights an important diversity of CBM98–CBM48–GH13_47-containing proteins to be functionally and structurally explored in the future.

To gain insight into conserved features of CBM98, all CBM98 sequences were aligned and those exhibiting over 90% redundancy were removed and used in the ConSurf server to map conserved residues onto BoGH13A$_{Sus}$ [81–83]. The most conserved residues are found in the core CBM98 β-sandwich (Fig. 8E, F). The CBM98 dual tryptophan-binding site is highly conserved, with the W92 equivalent sometimes replaced with a phenylalanine. BoCBM48 is nonetheless a bit distinct from other CBM48s that immediately follow a CBM98 for two reasons: (1) The loop between β8 and β9 in BoCBM48 (detailed more in the Discussion) that covers a predicted CBM48-binding site is ~ ten amino acids shorter than in most other proteins and (2) E274 that coordinates a Ca^{2+} ion is more often an aspartate than a glutamate in other CBM48s that immediately follow a CBM98. Furthermore, the neighboring residue in BoCBM48 (K275) hydrogen bonds with CBM98 ligand but in similar CBM48s, this residue is a proline (Supplementary Fig. 18F). Thus, while BoGH13A$_{Sus}$ is the founding member of both the GH13_47 subfamily and CBM98 family, the interplay between its CBM98 and CBM48 may differ from other similar enzymes.

Discussion

Multiple studies have examined the structure and function of various PUL from gut *Bacteroides* for the recognition and degradation of carbohydrates. Fewer have compared functional differences between mostly syntenic PUL [76, 87, 88]. Still fewer have compared the structural differences among functionally similar enzymes or lipoproteins between species [89]. Given the importance of characterizing the mechanistic underpinnings of functional diversity in the gut microbiome, we examined the structural differences between two surface α-amylases in *Bacteroides thetaiotaomicron* (Bt) and *Bacteroides ovatus* (Bo) that otherwise act as part of syntenic PUL required for starch degradation. Indeed, despite having very similar starch PUL (Fig. 1B), Bo outcompetes Bt in an in vitro co-culture grown on amylopectin [90].

A key structural difference between BoGH13A$_{Sus}$ and SusG is the position and type of CBMs found therein. More interestingly, the importance of these CBMs to overall catalysis and growth on starch is distinct. Our previous work with SusG demonstrated that the CBM58 is not necessary for activity or growth [23, 24]. This contrasts with BoGH13A$_{Sus}$ in which competent binding at CBM98 is absolutely required for growth on polysaccharides and enhances binding to and catalytic activity on these substrates (Figs. 6, 7D). Moreover, the inverted orientation of the enzymes with respect to the membrane was surprising. That the CBM98 is proximal to the membrane initially made us think that it was unlikely to participate in direct starch capture. For example, some CBMs mediate cell wall attachment in bacteria [91, 92]. However, the CBM* protein, which was mutated to abrogate CBM98 starch binding (Table 2, Supplementary Fig. 11), has significantly decreased starch binding and activity in vitro, suggesting a role in starch breakdown (Fig. 6).

CBM98 is always N-terminal to a CBM48. This domain architecture is akin to α-glucan debranching enzymes and pullulanases that target α1,6 bonds and feature a CBM48 N-terminal to a GH13 catalytic domain and in some cases, additional CBMs and unclassified domains [93]. BoGH13A$_{Sus}$ does not hydrolyze α1,6 bonds, however (Fig. 2). Nonetheless, the most structurally related CBM48 with bound ligand to BoCBM48 according to the DALI server is that from the *Escherichia coli* branching enzyme (Z score: 12.8; PDB ID: 4LQ1 [94]) at the N terminus of the protein [43]. Another closely related CBM48 from the branching enzyme in *Cyanothece* sp. ATCC 51142 (*Z*-score: 12.6; PDB ID: 5GQV [59]) was also included for comparison because its two binding sites are occupied. Indeed, while many CBM48s do not bind oligosaccharides (discussed below), those in branching enzymes tend to do so [93]. BoCBM48 has two additional β strands, β8 and β9, that are loops in other CBM48s (Supplementary Fig. 19A). A large loop (K275-P288) between these two strands occludes a potential binding site, centered at W219, that is conserved in the CBM48 family (Supplementary Fig. 19A). Similarly, an extended loop between β2 and β3 precludes binding at another enigmatic CBM48 site that is poorly conserved (Supplementary Fig. 19B).

Lack of CBM48 maltooligosaccharide binding is not uncommon—those appended to isoamylases, glycogen debranching enzymes and pullulanases lack conserved residues at canonical site 1 from CBM20s, which are evolutionarily related to CBM48s, but have a conserved tryptophan at site 2 [93, 95]. BoCBM48 also has this site 2 tryptophan and a similar site 1 to CBM20 (Supplementary Fig. 19C). It appears that BoCBM48, despite the conservation of some starch-binding residues, has diverged from the subset of CBM48s that bind starch within larger enzymes primarily due to extended loops that preclude binding. This is distinct from other CBM48s that lack a completely conserved canonical site 1. As CBM98 always co-occurs with a CBM48, it is possible that these modules co-evolved leading to the divergence of this CBM48 from other ones. BoCBM48 may function as a sort of linker to help position substrate between the CBM98-binding site and the active site for optimal catalysis. It may also be necessary for stability and structural integrity since attempts to express a CBM98-GH13 construct of BoGH13A$_{Sus}$ were unsuccessful.

Despite drastic differences in structure and function between their α-amylase enzymes, Bt and Bo resume growth on starch when complemented with the α-amylase from the other species. It is notable that complementation of the Bt enzyme in Bo was more efficient than the complementation of BoGH13A$_{Sus}$ into Bt (Fig. 7A). While we observed similar enzyme amounts produced in both cases (Supplementary Fig. 15), it is possible that small changes in cell surface enzyme concentration, differences in function or aspects of the other bacterium's outer membrane leaflet made function less than optimal.

With respect to function, the SusG and BoGH13A$_{Sus}$ active sites likely recognize different motifs in starch. Using acarbose as a proxy for an exclusively α1,4-linked maltooligosaccharide bound to the BoGH13A$_{Sus}$ active site, we compared the disposition of acarbose in BoGH13A$_{Sus}$ to G7 bound to SusG. SusG enforces a much more curved conformation onto the oligosaccharide, whereas BoGH13A$_{Sus}$ does not interact to the same extent with the non-reducing or reducing ends, leading to a more open chain structure (Supplementary Fig. 19D–F). What this means in the context of growth, however, is unclear until a more quantitative kinetic analysis on each enzyme is performed.

In our previous work, we utilized single molecule imaging to observe the movement of SusG on the cell surface [26–28]. Measurement of the SusG diffusion coefficient in various Sus protein deletion backgrounds has lead us to propose, for the BtSus, a model of dynamic assembly whereby Sus proteins associate and dissociate during starch catabolism [26, 27]. Thus far, an intact complex of all components has not been observed although this has been seen in levan- and dextran-targeting PUL in Bt [96]. Nonetheless, the dynamics of BoGH13A$_{Sus}$ association with BtSus proteins in the presence of starch may not be optimal for growth, but SusG association with BoSus is. BoSus protein dynamics at the cell surface have not been characterized but warrant further investigation given the major structural differences between SusG and BoGH13A$_{Sus}$.

Beyond the topological differences in these enzymes, we were surprised to find that BoGH13A$_{Sus}$ coordinates Mn^{2+} and that Mn^{2+} supplementation augmented growth of the BtΔ*susG* + *bovatus_03803* strain (Fig. 7B, Supplementary Fig. 16A). To our knowledge, this is the first example of a Mn^{2+}-dependent amylase, although there are

Mn^{2+}-dependent GH4 enzymes that act on α-glucosides [97, 98]. Dietary Mn^{2+} is absorbed in the small intestine [99]. Excess Mn^{2+} is returned to the digestive tract via the liver and is readily detected in the lumen of the murine colon [100, 101]. In agricultural applications, exogenous supplementation of feed enzymes such as xylanase and α-glucanase increases Mn^{2+} release from feedstuffs such as wheat, barley and corn, raising the possibility that microbial sources of these enzymes could promote additional deposition of Mn^{2+} and other metals in the gut [102, 103]. Still, most of the research around Mn^{2+} in the gut microbiota involves the interplay between host and pathogen in nutritional immunity [104]. Mn^{2+} is a cofactor in some bacterial superoxide dismutases that detoxify superoxide radicals generated by the immune response [105]. In turn, the host produces calprotectin in the intestinal lumen to sequester metals including Mn^{2+} and some enteric pathogens overcome this by upregulating cation transporters [106].

At this point, the influence of colonic Mn^{2+} levels on the gut microbiota is unknown, much less how it influences polysaccharide degradation [104]. This may be of interest given that Mn^{2+} binding could be a hallmark of the GH13_47 subfamily (Supplementary Fig. 19G, H) and this subfamily is highly represented in the Bacteroidetes (Fig. 8C and Supplementary Fig. 18 B–D).

Increasing starch flux to the gut, via the consumption of resistant starches that escape host enzymes or by inhibiting those enzymes with pharmaceuticals like acarbose, is of interest given the therapeutic benefits [107, 108]. However, individual responses to these interventions vary [107, 109], so knowing what bacterial starch degradation machinery a person's microbiota harbors will be key to prescribing the correct treatment. For example, GH13_47s and GH13_36s are not the only starch active enzymes that can support Bacteroidetes growth. Some species within this phylum grow on starch but do not encode a GH13_47 or GH13_36 [29, 110]. Broadly speaking, the structure and function of other Bacteroidetes α-amylases should be studied and specifically, CBM98-containing enzymes warrant further investigation given that they vary from BoGH13A$_{Sus}$.

Materials and methods

Bacterial genetic manipulation and growth conditions

Bacteroides thetaiotaomicron (VPI-5482) (Bt) and *Bacteroides ovatus* (ATCC 8483) (Bo) were routinely grown on tryptone–yeast extract–glucose (TYG) medium [111], minimal medium (MM) [12] or on brain heart infusion (Becton Dickinson) agar supplemented with 10% horse blood (Colorado Serum Co.). For all growths and genome manipulations, the parent strains BtΔ*tdk* and BoΔ*tdk* were utilized and are considered wild type as they contain no *sus* mutations [85, 112]. *bovatus_03803* (which encodes BoGH13A$_{Sus}$) was deleted from the Bo genome using a previously described counter selectable allelic exchange vector pExchange-*tdk* [112] to generate the strain BoΔ*bovatus_03803*. BtΔ*susG* was derived previously [24]. BoΔ*bovatus_03803* and BtΔ*susG* were used as recipient strains to introduce α-amylase from the opposite organism into the native α-amylase gene location of each *sus* locus. *bovatus_03803* was cloned into pExchange containing 750 bp flanks (*bt_3697* and *susF*), while *susG* was cloned into pExchange containing 750 bp flanks (*bovatus_03802* and *bovatus_03804*), both using overlap extension PCRs.

To assess a pathology in the BtΔ*susG* + *bovatus_03803* strain, *bovatus_03803* was cloned into a pNBU2 vector with a constitutively active promoter (sigma 70, *rpoD*) and complemented into one of two tRNA^{ser} sites [12, 75].

To generate Bo-CBM*, Bo-surf* and Bo-CBM*surf* strains, a pExchange vector containing 750 bp flanks on either side of the wild-typed cloned *bovatus_03803* gene was generated. This was subsequently used as a template for mutagenesis (described below) to iteratively generate each of the variants pExchange vectors. BoΔ*bovatus_03803* was used as the recipient strain to facilitate faster screening of potential positive clones.

All manipulated Bt and Bo strains were screened for proper incorporation of the target gene by PCR-amplifying the expected inserts and subjecting them to Sanger sequencing at the University of Michigan DNA sequencing core or Azenta Life Sciences. All strains and constructs are listed in Supplementary Table 1, and primers are listed in Supplementary Table 2.

For plate reader growths, strains were inoculated into TYG from freezer stocks and grown overnight (16 h) at 37 °C in a Coy anaerobic chamber (85% N_2/10% H_2/5% CO_2). Cells were then back diluted 1:50 into MM containing 5 mg/mL glucose and grown overnight. The next day, cells were centrifuged and washed with 2× MM containing no carbon source and diluted 1:100 into MM with 2.5 mg/mL of experimental carbohydrate ± 0.5 mg/ml maltose in parallel with a MM + 2.5 mg/ml glucose control. Substrates included maltose (Sigma), potato amylopectin (Sigma), potato starch (Sigma), maltoheptaose (Carbosynth), pullulan (Megazyme) and bovine liver glycogen (Sigma). Kinetic growth experiments were performed in 96-well plates in an anaerobic chamber at 37 °C outfitted with a BioTek Biostack plate-handler and Powerwave HT plate reader. An OD_{600} was recorded every 10 min with the average of 3 replicates.

Gene cloning and site-directed mutagenesis for heterologous protein expression

The gene encoding BoGH13A$_{Sus}$ (residues 21–758) from Bo (*bovatus_03803* [GenBank: ALJ48408], previous annotation: *bacova_03514* [GenBank: EDO10881]) was amplified from genomic DNA purified from *B. ovatus* ATCC 8483 using PCR. Amplification was done using an N-terminal primer encoding a tobacco etch virus (TEV) protease cleavage site between the 6x-Histidine tag and start of the BoGH13A$_{Sus}$ coding sequence. PCR was performed using the Phusion Flash High-Fidelity system. The amplicon was inserted into the pETite vector using the Expresso T7 Cloning system from Lucigen. The BoGH13A$_{Sus}$-pETite vector was used as a template for subsequent mutagenesis reactions and PCRs. All constructs were cloned with the same 6x-His-TEV site at the N terminus as BoGH13A$_{Sus}$-pETite. CBM98-pETite contains residues 22–143. CBM98–CBM48–pETite contains residues 22–296. CBM98binding residues W92 and W98 were mutated to alanine residues in both a CBM98 and BoGH13A$_{Sus}$ background to generate CBM98–W92,98A and BoGH13A$_{Sus}$-CBM*, respectively. Surface site binding site residues W555 and Y592 were mutated to alanine residues in a BoGH13A$_{Sus}$ background to generate BoGH13A$_{Sus}$ surf*. This construct was used as template to mutate W92 and W98 to alanines to make a combined CBM98 and surface binding site mutant, BoGH13A$_{Sus}$-CBM*surf*. Putative catalytic amino acids D581 and D477 were mutated to asparagine residues, while E523 was mutated to glutamine within BoGH13A$_{Sus}$. Mutagenesis reactions proceeded as follows: 150 or 300 ng of template DNA was used in 50 µl reactions containing 0.5 µM each of forward and reverse primers. PCR cleanup was followed by DpnI digest to remove template DNA and subsequent transformation of 10G *E. coli* competent cells with reaction product. Mutations were confirmed with Sanger sequencing at the University of Michigan DNA Sequencing Core or Azenta Life Sciences. All constructs are listed in Supplementary Table 1, and primers are listed in Supplementary Table 2.

Protein expression and purification

Rosetta (DE3) pLysS cells were transformed with the desired BoGH13A$_{Sus}$ construct. 20 ml of an overnight culture from LB medium were used to inoculate 1 L of TB medium the following morning. Cells were grown to an OD_{600} of 0.8–1 at 37 °C with shaking (180 RPM) and subsequently cooled on ice water. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cells were grown at 22 °C for 18 h. The next day, cells were harvested by centrifugation for 10 min at 10,000×*g*. Cell pellets were stored at -80 °C.

All constructs were purified on a 5-mL His-Trap metal-affinity column (GE Healthcare and Cytiva Life Sciences). The column was equilibrated in lysis buffer (25 mM NaH_2PO_4, 500 mM NaCl, 20 mM imidazole, pH 7.5). Cells were resuspended in lysis buffer, sonicated on ice and then centrifuged at 4 °C for 30 min at 30,000×*g*. Cleared lysate was applied to the column at room temperature, which was then washed with lysis buffer until reaching a baseline A_{280}. Protein was eluted with a 70 mL gradient from 0 to 100% elution buffer (25 mM NaH_2PO_4, 500 mM NaCl, 300 mM imidazole, pH 7.5). Fractions containing target protein were confirmed via SDS-PAGE and pooled. Recombinant TEV protease was incubated with protein at a 1:30 molar ratio in lysis buffer to remove the 6x-His tag. The reaction was allowed to proceed either overnight at room temperature or for 2 days at 4 °C. The mixture was applied to a His-Trap column to remove the His-tagged TEV protease, 6x-His tag and undigested SusG protein. Pure fractions were confirmed via SDS–PAGE and dialyzed against 20 mM HEPES (pH 7.0) and 100 mM NaCl at 4 °C overnight.

Crystallization and data collection

Native structure determination

BoGH13A$_{Sus}$ was concentrated to 10 mg/mL based on an estimated extinction coefficient of 163,320 M^{-1} cm^{-1}. Crystallization conditions were initially surveyed via sitting drop vapor diffusion at room temperature using an Art Robbins Gryphon LCP crystallization robot. Screening was carried out using commercially available kits from Morpheus, Hampton and Jena Bioscience. Hanging drop refinement was pursued using conditions from the Morpheus screen including their 1.2 M ethylene glycols and precipitant 4 (25% v/v methyl-2,4-pentanediol [MPD]; 25% poly(ethylene glycol) [PEG] 1000; 25% w/v PEG 3350 for a total of 75% precipitant). 1 M solutions of 2-(*N*-Morpholino)-ethane sulfonic acid (MES) and imidazole were prepared (Sigma). Imidazole was added drop wise to 30 mL of MES to reach a pH of 6.5. Optimization of crystallization conditions was subsequently carried out using 0.1–0.13 M ethylene glycols and 34–36% precipitant mix 4 in 100 mM imidazole/MES buffer, pH 6.5 using 1 µl of well solution and 2 µl of protein in a hanging drop format with 500 µl of well solution. Large, single crystals were transferred from 34% precipitant, 0.1 M ethylene glycols and placed into a cryoprotectant solution of 80% well solution and 20% ethylene glycol for 30 s before flash-freezing in liquid nitrogen. These crystals belonged to the monoclinic space group $P2_1$ with unit cell dimensions of $a = 100.3$ Å, $b = 148.3$ Å, $c = 112.7$ Å, $\beta = 91.01°$. X-ray data were collected at the Advanced Photon Source Life Science Collaborative Access Team (LS-CAT) beamline ID-F at Argonne National Laboratories in Illinois. Data

were processed and scaled in Xia2 [113] with XDS [114]. Molecular replacement was carried out in Balbes [115] and the model selected was PDB 3M07, a structural genomics deposition of a putative α-amylase from *Salmonella typhimurium* str. LT2. Several rounds of Autobuild in Phenix [116] produced a model containing four monomers in the asymmetric unit. The model was manually adjusted in Coot [117, 120] followed by refinement in Refmac [121]. Data were re-processed in Xia2 using DIALS [122] and solved by molecular replacement using Phaser [123] in Phenix [124, 125] with the initial BoGH13A$_{Sus}$ model refined in Coot. The final model was produced via manual adjustment in Coot and refinement in Refmac.

BoGH13A$_{Sus}$–acarbose structure determination

9.8 mg/mL BoGH13A$_{Sus}$ with 10 mM acarbose (Sigma) was screened for crystallization as previously described with refinement pursued using the Morpheus monosaccharides mix and hand-made mixes of their buffer system 2 (1 M sodium 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid [HEPES]/3-morpholinopropane-1-sulfonic acid [MOPS], pH 7.5) and precipitant mix 2 (40% v/v PEG methyl ether [MME] 550; 20% w/v PEG 20,000). Optimized crystallization conditions included 0.12 M monosaccharides mix, 28–32% precipitant mix 2 and 100 mM HEPES/ MOPS pH 7.5 using 1 µl of well solution and 2 µl of protein in hanging drop format with 500 µl of well solution. Suitable crystals were transferred from 30% precipitant to a cryoprotectant solution comprised of 70% well solution, 20% glycerol and 10% acarbose (final concentration of 10 mM) for 30 s and flash frozen in liquid nitrogen. These crystals also belonged to the space group $P2_1$ with unit cell dimensions of $a = 100.1$ Å, $b = 125.3$ Å, $c = 150.8$ Å, $\beta = 102.1°$.

X-ray data were collected at the Advanced Photon Source Life Science Collaborative Access Team (LS-CAT) beamline ID-D at Argonne National Laboratories in Illinois. Data were processed with Xia2 [113] and scaled in Aimless [126]. The native BoGH13A$_{Sus}$ structure was used as the molecular replacement model in Phaser [123] within Phenix [127] followed by Autobuild [116] to derive the initial model and map for refinement. The asymmetric unit was comprised of four monomers. The model was manually adjusted in Coot [117, 120] followed by refinement in Refmac [121]. Carbohydrate geometry was validated in Privateer [128].

BoGH13A$_{Sus}$–maltoheptaose structure determination

All three active site variants (D477N, E523Q and D581N) were screened for crystallization properties with 9.8 mg/ml protein as described above with 10 mM maltoheptaose (Carboexpert), but the E523Q protein yielded the most workable crystals. Crystals were further refined out of Morpheus screening conditions using hanging drop vapor diffusion. Ethylene glycols, precipitant mix 4 (described above), and buffer system 3 (created by mixing 1 M 2-Amino-2-hydroxymethyl-propane-1,3-diol [Tris-base] and 1 M *N*,*N*-bis(2-hydroxyethyl)glycine [Bicine] to a pH of 8.5) were used in refinement. Suitable crystals grew in hanging drop format with 500 µl of well solution from drops containing 2 µl of protein and 1 µl of a well solution comprised of 36% precipitant mix 4, 0.08 M ethylene glycols mix and 0.1 M buffer system 3. Crystals were moved to a cryoprotectant solution containing 73% well solution, 18% ethylene glycol and 9% maltoheptaose (final concentration of 9.1 mM) for 30 s and flash frozen in liquid nitrogen. Crystals belonged to the space group $P2_1$ with unit cell dimensions of $a = 98.7$ Å, $b = 128.6$ Å, $c = 149.8$ Å, $\beta = 105.4°$.

X-ray data were collected at the Advanced Photon Source Life Science Collaborative Access Team (LS-CAT) beamline ID-D at Argonne National Laboratories in Illinois. Data were processed and scaled as described for the acarbose-bound structure. Phasing, initial model generation and refinement were also done as described above for the acarbose-bound structure. Carbohydrate geometry was validated in Privateer [128].

Thin-layer chromatography (TLC)

0.5 µM BoGH13A$_{Sus}$ was incubated at 37 °C overnight in 20 mM HEPES, 100 mM NaCl, pH 7.0 with 5 mg/ml of one of the following polysaccharides or oligosaccharides: maltose (Sigma), maltotriose (Carboexpert), maltotetraose (Carboexpert), maltopentaose (Carbosynth), maltohexaose (Carbosynth), maltoheptaose (Carboexpert), 6^3-α-D-glucosyl-maltotriose (Megazyme), 6^3-α-D-glucosyl-maltotriosyl-maltotriose (Megazyme), potato amylopectin (Sigma), bovine liver glycogen (Sigma), pullulan (Megazyme), dextran (Sigma), acarbose (Sigma), α-cyclodextrin (Sigma), β-cyclodextrin (Sigma), panose (Sigma) and isomaltose (Sigma). Two microliters of each reaction was spotted onto a 20 × 20 cm silica gel 60G F_{254} glass plate (MilliporeSigma). Controls of each sugar without BoGH13A$_{Sus}$ were spotted. The plate was allowed to air-dry and then placed into a solvent chamber with 1-butanol/acetic acid/water in a 2:1:1 v/v ratio. One irrigation was performed after which the plate was dried with a hair dryer. The plate was then immersed in a developer solution containing sulfuric acid/ethanol/water in a 3.2:75.3:21.5 ratio with 1% orcinol (Sigma) for 30 s. The plate was dried with a hair dryer and visualized after heating in an oven for 10–20 s.

BoGH13A$_{Sus}$ is annotated as a malto-oligosyltrehalose trehalohydrolase, TreZ (EC 3.2.1.141). These enzymes hydrolyze α1,4-glycosidic linkages between trehalose and a glucan comprised of α1,4 linkages. It was not possible to obtain suitable substrate to test the forward reaction, so

$BoGH13A_{Sus}$ was incubated with trehalose and maltose to see if the reverse reaction would occur. No maltosyl trehalose was synthesized (data not shown).

In separate experiments, 10 mg/ml WT and E523Q $BoGH13A_{Sus}$ and 1 µM WT and $BoGH13A_{Sus}$ E523Q were incubated overnight at room temperature with 10 mM G7 in 100 mM Bis–Tris/Bicine pH 8.5, 0.08 Morpheus ethylene glycols mix and 36% Morpheus precipitant mix 4. The reaction was diluted twofold, and 3 µl was spotted and resolved on TLC as described above. 1.25 mg/ml of each of G1–G7 was diluted into the same reaction buffer to run as standards (3 µl).

To observe possible $BoGH13A_{Sus}$ transglycosylation of acarbose in crystals, hanging drops containing crystals that formed in the same conditions as described for the acarbose-bound structure determination were smashed to dissolve the crystals. Three microliters of the smashed mixture was loaded onto a TLC plate, along with the same volume of 1.25 mg/ml acarbose and G1–G7 as standards. Standards were diluted in water. The TLC plate was visualized as described above.

Isothermal titration calorimetry (ITC)

CBM98 and CBM98–CBM48 binding to oligosaccharides was assessed via isothermal titration calorimetry (ITC) at 25 °C using a TA instruments low-volume NanoITC, except where noted. All constructs (CBM98, CBM98–W92,98A, CBM98–CBM48, CBM98–CBM48–W92,98A) were used at 100 µM. 2 mM maltoheptaose, α-cyclodextrin and β-cyclodextrin were titrated, while maltotriose was titrated at 10 mM. Finally, 0.5% autoclaved potato amylopectin was also titrated to assess binding. Exceptions to the above include the following: CBM98–CBM48 binding to α- and β-cyclodextrin was conducted in a standard volume instrument using 25 µM protein and 1 mM ligand with a ligand into buffer blank used to subtract the heat of dilution. A constant blank correction was used for oligosaccharides while an amylopectin blank into dialysis buffer was used to subtract the heat of dilution for those experiments. All data were analyzed using the manufacturer's NanoAnalyze software. An independent binding model was used for analysis except for maltotriose. The binding in this case was low, which was difficult to model, so N was set to 1. Finally, the concentration of amylopectin was manually adjusted to achieve an N of 1 according to [73].

Native polyacrylamide gel electrophoresis (PAGE)

Native PAGE was used to assess $BoGH13A_{Sus}$ binding to potato amylopectin, glycogen, pullulan and autoclaved potato starch [73, 74]. Six percent native polyacrylamide gels in 0.375 M Tris–HCl, pH 8.8, with or without the addition of 0.1% carbohydrate were subjected to 75 V for 6 h. The gel rigs were placed in ice water for the run to prevent gels from overheating. Five microliters of 0.5 mg/ml $BoGH13A_{Sus}$ variant or BSA (negative control) was loaded per well. Gels were stained for 30 min in 0.1% Coomassie R-250 in 10% acetic acid, 50% methanol and 40% water. They were de-stained for 30 min in the same solution lacking the Coomassie, put into fresh de-stain for 30 min, stored in 5% acetic acid overnight and visualized the next day.

Wild-type and variant $BoGH13A_{Sus}$ activity on an oligosaccharide

Wild-type and variant $BoGH13A_{Sus}$ activity on an oligosaccharide was assessed using a colorimetric amylase activity kit from LSBio (LS-K10, Seattle, WA). Briefly, ethylidene-para-nitrophenol-maltoheptaose (blocked pNP-G7) is hydrolyzed by $BoGH13A_{Sus}$ and exposes a new non-reducing end. This non-reducing end becomes a substrate for an α-glucosidase in the kit that causes the eventual release of 4-nitrophenol (pNP) from $BoGH13A_{Sus}$ products. The manufacturer's instructions were followed using 150 nM of WT, CBM*, surf*, CBM*surf* and E523Q $BoGH13A_{Sus}$ at 25 ºC in duplicate. The absorbance was read at 405 nm in a Biotek SynergyH1 plate reader every minute upon addition of enzyme. An initial rate was obtained from the linear portion of the curves from six to twenty minutes. pNP concentrations were derived using a standard curve of pNP in reaction buffer conditions.

Wild-type and variant $BoGH13A_{Sus}$ activity on polysaccharides

The dinitrosalisylic acid-based method for the detection of reducing ends [129] was followed according to [130]. A stock reagent containing 1% 3,5-dinitrosalisylic acid (Sigma), 1% sodium hydroxide (Fisher Scientific), 0.2% phenol (Sigma) and 0.5% sodium sulfite (Sigma) was prepared. Working reagent ($DNSA_{20}$) was made by adding 1 µl of 20% α-glucose (Sigma) per ml of DNSA stock reagent for a final glucose concentration of 0.02%. A standard curve from 0 to 8 mM maltose was generated each time new working reagent was prepped. The maltose was diluted in reaction buffer and an equivalent volume was processed as for the reaction conditions (see below).

To compare WT and variant $BoGH13A_{Sus}$ (CBM*, surf* and CBM*surf*) activity on different polysaccharides, 200 µl reactions were set up in duplicate. 100 µl of 2×buffer and 2×protein were added to 100 µl of 2×carbohydrate for a final concentration of 100 nM protein, 20 mM HEPES pH7, 100 mM NaCl and 2.5 mg/ml carbohydrate (potato amylopectin, pullulan, glycogen or autoclaved potato starch). A no enzyme control was used for each polysaccharide to

obtain the time = 0 reading. Reactions were carried out at 25 °C. After ten minutes, 150 µl of each reaction was added to 150 µl of DNSA working reagent. The mixture was transferred to Eppendorf tubes and boiled at 100 °C in a heat block for 15 min, cooled in ice water for 5 min and equilibrated to room temperature. Two 100 µl samples were taken from each replicate and standard curve sample and transferred to a clear 96-well plate (Corning) to be read in a SynergyH1 plate reader at 540 nm. The total concentration of reducing ends was calculated according to the standard curve by subtracting the no enzyme control reading from the ten-minute timepoint.

Red starch activity assay in the presence of various metals

A stock solution of red starch (Megazyme) was prepared in KCl according to the manufacturer's instructions. 1.5 mg/ml and 15 mg/ml red starch and 750 nM BoGH13A$_{Sus}$ with no added metals or 15 µM, 150 µM or 1.5 mM $CaCl_2$ and either $MgCl_2$, $MnCl_2$, $CuSO_4$, $CoCl_2$, $ZnCl_2$, $NiSO_4$ or additional $CaCl_2$ were incubated at 37 ºC for 5 min in 2-ml Nunc deep well plates (Thermo). One hundred and twenty microliters of protein solution was added to 60 µl of red starch to achieve final concentrations of 500 nM enzyme, 0.5 or 5 mg/ml starch and 20 µM, 200 µM or 2 mM total added metal in 20 mM HEPES, 100 mM NaCl pH 7.0. The reaction was allowed to proceed for 5 min at 37 ºC and stopped with the addition of 300 µl of 95% ethanol. Mixtures were transferred to an Eppendorf tube to spin for 20 min at 1,000 x*g*. 2 × 200 µl of each sample was transferred to a clear 96-well plate and read in a SynergyH1 plate reader at 510 nm.

Sequence alignments and phylogenetic analyses

SusDEF (BT_3701, BT_3700 and BT_3699, GenBank: AAO78806.1, AAO78805.1 and AAO78804.1, respectively) were aligned with the corresponding proteins from Bo, BoSusD, BoSGBP$_{Sus}$-B and BoSGBP$_{Sus}$-A, respectively (Bovatus_03806, Bovatus_03805 and Bovatus_03804, GenBank: ALJ48411.1, ALJ48410.1 and ALJ48409.1, respectively) using the ClustalOmega webserver https://www.ebi.ac.uk/Tools/msa/clustalo/ using default parameters [131, 132]. Binding site residues from the Bt proteins and the corresponding Bo protein residue are boxed according to whether or not they are conserved, if experimental evidence exists that mutating the residue leads to loss of binding competency and if the residue is in close proximity to bound ligand in the Bt protein structures [24, 112, 133].

SusG and BoGH13A$_{Sus}$ alignments were carried out in MegAlign Pro within DNASTAR with the Global: Needleman–Wunsch algorithm using default parameters [134, 135]. A BLOSUM40 substitution matrix was used to align full-length SusG with full-length BoGH13A$_{Sus}$. A BLOSUM55 matrix was used to align the GH13 portions of SusG and BoGH13A$_{Sus}$. The BLOSUM matrices were chosen after initial alignments with a range of matrices and subsequently 40 or 55 were chosen as they gave the highest scores.

Selected GH13 sequences from subfamilies 36, 23, 31, 10, 47 and group_re1-3 (see Supplementary Table 3) were submitted to the NGPhylogeny.fr workflow using standard parameters [136]. MAFFT was used to create the initial alignment [137]. The alignments were trimmed and cleaned using BMGE [138]. FastME was used to construct a distance matrix and output tree, which was ultimately rendered using a Newick display [136, 139, 140]. The tree was displayed with iTOL [141].

Selected CBM98–CBM48–GH13_47 enzymes (see Supplementary Table 3) were aligned with ClustalOmega within MegAlign Pro to compare the CBM48s and Mn^{2+}-binding sites within these proteins [131, 135].

CBM20 from *Aspergillus* niger (residues 509–616, [142]), CBM48 from barley limit dextrinase (residues 126–226, [143]) and BoCBM48 (residues 173–295) were aligned using MAFFT within MegAlign Pro with default parameters [135]

Full-length CBM98 domain sequences were extracted from the CAZy database and aligned using MAFFT (v7.453) with default parameters [29]. A total of 812 sequences were aligned. Sequences were culled in JalView (v2.11.2.6) at a 90% redundancy cutoff to generate an appropriate file size to submit to the Consurf server [78–80]. There were 288 sequences in the final alignment.

Native and inductively coupled plasma mass spectrometry

Native mass spectrometry data were collected on a QE Exactive Ultra-High Mass Range (UHMR) orbitrap system (ThermoFisher). Tune settings were carefully selected such that the protein would remain intact for the duration of the experiment. Prior to MS, samples were buffer exchanged into 200 mM ammonium acetate (pH = 8) (Sigma) to a final concentration of 10 µM protein using P-6 Biospin Columns (Bio-Rad). Ions were generated using nanoelectrospray ionization in positive polarity through direct infusion with the nanospray FlexIon source (ThermoFisher). 1.5 kV of capillary voltage was applied to a borosilicate emitter that was coated in gold and fabricated in house using a P-97 pipette puller (Sutter Instruments). The capillary temperature was set to 275°, and the S-lens RF was set to 80. Data were collected with a resolution of 200,000 at 200 Th and were visualized with FreeStyle (ThermoFisher) and deconvoluted with UniDec [144].

Inductively coupled plasma (ICP) mass spectrometry data were collected with a Nexion 2000 ICP-MS (PerkinElmer)

that utilizes argon plasma. 10 µM $BoGH13A_{Sus}$ samples were diluted with 2% nitric acid (Sigma) prior to analysis and ionization. Internal standards (PerkinElmer), namely bismuth, holmium, indium, lithium-6, scandium, terbium and yttrium, were employed to ensure data reproducibility. Calibration curves were constructed for ^{55}Mn and ^{44}Ca, ranging from 1 to 50 ppb via ICP-MS, with a blank sample also being run. ICP-MS data for $BoGH13A_{Sus}$ samples were collected in triplicate by autosampling. Data were analyzed with Synergistix (PerkinElmer) and Excel (Microsoft).

Histogram and Venn diagram construction

Proteins containing either GH13_36 or GH13_47 modules were extracted from CAZy data in October 2022, corresponding to 614 and 486 proteins [29]. Histograms and Venn diagrams were generated using home-made Python (v3.8.10) scripts as follows.

Each protein was associated with additional information extracted from the private CAZy and PULDB data [79], using "mysql-connector-python" (v8.0.31) to import data and "pandas" (v1.2.3) to store in *DataFrames*. This information included: (i) the sequence length; (ii) the organism full-name (including its strain); (iii) the taxonomic phylum, to distinguish Bacteroidetes; (iv) the PUL identifier, if belonging to a predicted PUL in PULDB; (v) co-presence in this PUL of both SusA-like (GH13_46 subfamily [145]) and SusB-like (GH97 private subfamily 2 in CAZy) homologs, or not.

To compare the protein lengths between these two subfamilies, we generated histograms using "seaborn" (v0.12.2). The impact of extreme values was reduced by applying a log2 transformation.

Venn diagrams were produced using "matplotlib-venn" (v0.11.7) based on two distinct features: organism names (orange/blue diagrams) and presence in a PUL (pink/purple diagrams).

Western blot analysis

Bo and Bt strains were inoculated from freezer stocks in TYG and back diluted 1:50 into MM + 5 mg/ml glucose. The next day, strains were back diluted into 5 ml of MM + 5 mg/ml maltose to achieve a starting OD_{600} of 0.15 – 0.2. Strains were grown anaerobically at 37 °C to an OD_{600} of 0.65–0.72 after which the cells were pelleted. The pellet was resuspended in 50 µl of leftover media after centrifugation and mixed with 50 µl of 2 × sample buffer containing β-mercaptoethanol and extra SDS. Samples were boiled for 10 min at 100 °C, and 5 µl of total cell lysate or 2.5 ng of boiled purified protein was loaded in an 8% SDS–PAGE gel. Gels were run at 180 V for 65 min and protein transferred to an Immobilon PVDF membrane using the Bio-Rad Trans-Blot turbo system according to the manufacturer's instructions. Membranes were blocked for 1 h at room temperature or overnight at 4 °C in 20 ml of Tris-buffered saline supplemented with 0.5% Tween-20 (TBS-T), 2% normal goat serum (Jackson Immunoresearch) and 0.02% NaN_3 (blocking buffer). Membranes were rinsed with TBS-T before incubating with 20 ml of a 1:500 dilution of αBoGH13A_{Sus} (Lampire Biological Laboratories) or αSusG [24] in blocking buffer for 1 h at room temperature. Membranes were washed 4 × 5 min in TBS-T before incubation with 20 ml of a 1:3000 dilution of goat anti-rabbit HRP conjugate (Bio-Rad) for 1 h at room temperature. Membranes were washed 4 × 5 min in TBS-T before visualizing with luminol/peroxide (ThermoScientific) in a Syngene Pxi6 instrument using GeneSys software.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00018-023-04812-w.

Acknowledgements We thank members of the Koropatkin, Martens and Ruotolo laboratories for helpful feedback on this work.

Author contributions HAB cloned, expressed and purified protein constructs as well as performed crystallography experiments, X-ray data collection and structural model refinement, cloned and generated all variant Bt and Bo strains and performed bacterial growth experiments, TLC, ITC, Native PAGE, Western blot, enzyme kinetics, protein sequence analysis, and wrote and edited the original manuscript. ALD cloned, expressed and purified protein constructs, performed TLC and crystallography experiments and did structural model refinement, as well as collected ITC data. BRJ collected and analyzed native MS and ICP-MS data. ALP performed ITC experiments. MB performed bioinformatics analyses of CBM98 and GH13_47. REB assisted with ICP-MS data collection. ZW collected and processed X-ray data. BTR helped direct research and analyzed MS data. NT performed bioinformatics analyses of CBM98 and GH13_47, created the CBM98 family and GH13_47 subfamily within CAZy and analyzed their distribution in the private CAZy database. HAB, BLJ, MB and NT were responsible for data visualization. NMK helped design and direct research and wrote the original manuscript. All authors provided input on the manuscript.

Funding This work was supported by the following National Institutes of Health (NIH) grants from the National Institute of General Medical Sciences (NIGMS): R01-GM118475 to N.M.K. and R01-GM095832 to B.T.R. H.A.B was supported by an NIH Ruth L. Kirschstein Postdoctoral National Research Service Award F32-AT011278 from the National Center for Complementary and Integrative Health. The QE UMHR work was supported by the University of Michigan Biosciences Initiative. Use of the Pilatus 3 1 M detector was provided by grant 1S10OD018090-01 from NIGMS at the NIH. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant no.: 085P1000817). N.T. and M.B. were supported by the *Agence National de la Recherche* [grant number ANR-20-CE20-0022].

Availability of data and material $BoGH13A_{Sus}$ native, maltoheptaose and acarbose-bound structures were deposited with the RCSB Protein Data Bank, with PDB IDs of: 8DGE, 8DL1 and 8DL2, respectively. All plasmids, proteins, bacterial strains and other reagents generated in this study will be made freely available to any researcher wishing to use them for non-commercial purposes.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate. Not applicable.

Consent for publication Not applicable.

References

1. Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI (2012) Going viral: next generation sequencing applied to human gut phage populations. Nat Rev Microbiol 10:607–617. https://doi.org/10.1038/nrmicro2853
2. Kapitan M, Niemiec MJ, Steimle A, Frick JS, Jacobsen ID (2019) Fungi as part of the microbiota and interactions with intestinal bacteria. Curr Top Microbiol Immunol 422:265–301. https://doi.org/10.1007/82_2018_117
3. Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C, Schmitz RA (2018) Archaea are interactive components of complex microbiomes. Trends Microbiol 26:70–85. https://doi.org/10.1016/j.tim.2017.07.004
4. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC (2019) New insights from uncultivated genomes of the global human gut microbiome. Nature 568:505–510. https://doi.org/10.1038/s41586-019-1058-x
5. Wilson AS, Koller KR, Ramaboli MC, Nesengani LT, Ocvirk S, Chen C, Flanagan CA, Sapp FR, Merritt ZT, Bhatti F, Thomas TK, O'Keefe SJD (2020) Diet and the human gut microbiome: an international review. Dig Dis Sci 65:723–740. https://doi.org/10.1007/s10620-020-06112-w
6. Hansson GC (2020) Mucins and the microbiome. Annu Rev Biochem 89:769–793. https://doi.org/10.1146/annurev-biochem-011520-105053
7. Koh A, DeVadder F, Kovatcheva-Datchary P, Bäckhed F (2016) From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. Cell 165:1332–1345. https://doi.org/10.1016/j.cell.2016.05.041
8. Makki K, Deehan EC, Walter J, Bäckhed F (2018) The impact of dietary fiber on gut microbiota in host health and disease. Cell Host Microbe 23:705–715. https://doi.org/10.1016/j.chom.2018.05.012
9. McNeil NI (1984) The contribution of the large intestine to energy supplies in man. Am J Clin Nutr 39:338–342. https://doi.org/10.1093/ajcn/39.2.338
10. Cockburn DW, Koropatkin NM (2016) Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. J Mol Biol 428:3230–3252. https://doi.org/10.1016/j.jmb.2016.06.021
11. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol 9:e1001221. https://doi.org/10.1371/journal.pbio.1001221
12. Martens EC, Chiang HC, Gordon JI (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4:447–457. https://doi.org/10.1016/j.chom.2008.09.007
13. McNulty NP, Wu M, Erickson AR, Pan C, Erickson BK, Martens EC, Pudlo NA, Muegge BD, Henrissat B, Hettich RL, Gordon JI (2013) Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome. PLoS Biol 11:e1001637. https://doi.org/10.1371/journal.pbio.1001637
14. Brown HA, Koropatkin NM (2021) Host glycan utilization within the Bacteroidetes Sus-like paradigm. Glycobiology 31:697–706. https://doi.org/10.1093/glycob/cwaa054
15. McKee LS, La Rosa SL, Westereng B, Eijsink VG, Pope PB, Larsbrink J (2021) Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. Environ Microbiol Rep 13:559–581. https://doi.org/10.1111/1758-2229.12980
16. Martens EC, Koropatkin NM, Smith TJ, Gordon JI (2009) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J Biol Chem 284:24673–24677. https://doi.org/10.1074/jbc.R109.022848
17. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H (2017) Polysaccharide utilization loci: fueling microbial communities. J Bacteriol 199:e00860-e916. https://doi.org/10.1128/JB.00860-16
18. Cho KH, Salyers AA (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. J Bacteriol 183:7224–7230. https://doi.org/10.1128/JB.183.24.7224-7230.2001
19. D'Elia JN, Salyers AA (1996) Contribution of a neopullulanase, a pullulanase, and an α-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. J Bacteriol 178:7173–7179. https://doi.org/10.1128/jb.178.24.7173-7179.1996
20. Foley MH, Cockburn DW, Koropatkin NM (2016) The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. Cell Mol Life Sci. 73:2603–2617. https://doi.org/10.1007/s00018-016-2242-x
21. Reeves AR, Wang GR, Salyers AA (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. J Bacteriol 179:643–649. https://doi.org/10.1128/jb.179.3.643-649.1997a
22. Shipman JA, Cho KH, Siegel HA, Salyers AA (1999) Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. J Bacteriol 181:7206–7211. https://doi.org/10.1128/JB.181.23.7206-7211.1999
23. Koropatkin NM, Smith TJ (2010) SusG: a unique cell-membrane-associated α-amylase from a prominent human gut symbiont targets complex starch molecules. Structure 18:200–215. https://doi.org/10.1016/j.str.2009.12.010
24. Cameron EA, Kwiatkowski KJ, Lee BH, Hamaker BR, Koropatkin NM, Martens EC (2014) Multifunctional nutrient-binding proteins adapt human symbiotic bacteria for glycan competition in the gut by separately promoting enhanced sensing and catalysis. MBio 5:e01441-e1514. https://doi.org/10.1128/mBio.01441-14
25. Foley MH, Martens EC, Koropatkin NM (2018) SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*. Mol Microbiol 108:551–566. https://doi.org/10.1111/mmi.13949
26. Tuson HH, Foley MH, Koropatkin NM, Biteen JS (2018) The starch utilization system assembles around stationary starch-binding proteins. Biophys J 14:242–250. https://doi.org/10.1016/j.bpj.2017.12.015
27. Karunatilaka KS, Cameron EA, Martens EC, Koropatkin NM, Biteen JS (2014) Superresolution imaging captures carbohydrate utilization dynamics in human gut symbionts. MBio 5:e02172-14. https://doi.org/10.1128/mBio.02172-14
28. Geffroy L, Brown HA, DeVeaux AL, Koropatkin NM, Biteen JS (2022) Single-molecule dynamics of surface lipoproteins in *Bacteroides* indicate similarities and cooperativity. Biophys J 121:4644–4655. https://doi.org/10.1016/j.bpj.2022.10.024
29. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N (2022) The carbohydrate-active enzyme database:

functions and literature. Nucleic Acids Res 50:D571–D577. https://doi.org/10.1093/nar/gkab1045
30. Flint HJ, Whitehead TR, Martin JC, Gasparic A (1997) Interrupted catalytic domain structures in xylanases from two distantly related strains of *Prevotella ruminicola*. Biochem Biophys Acta 1337:161–165. https://doi.org/10.1016/s0167-4838(96)00213-0
31. Zhang M, Chekan JR, Dodd D, Hong PY, Radlinski L, Revindran V, Nair SK, Mackie RI, Cann I (2014) Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. Proc Natl Acad Sci U S A 111:E3708–E3717. https://doi.org/10.1073/pnas.1406156111
32. Machovič M, Svensson B, MacGregor EA, Janeček Š (2005) A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. FEBS J 272:5497–5513. https://doi.org/10.1111/j.1742-4658.2005.04942.x
33. Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B (2019) Bacteroidetes use thousands of enzyme combinations to break down glycans. Nat Commun 10:2043. https://doi.org/10.1038/s41467-019-10068-5
34. MacGregor EA, Janeček Š, Svensson B (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim Biophys Acta 1546:1–20. https://doi.org/10.1016/s0167-4838(00)00302-2
35. Arnal G, Cockburn DW, Brumer H, Koropatkin NM (2018) Structural basis for the flexible recognition of α-glucan substrates by *Bacteroides thetaiotaomicron* SusG. Protein Sci 27:1093–1101. https://doi.org/10.1002/pro.3410
36. Davies GJ, Wilson KS, Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem J. https://doi.org/10.1042/bj3210557
37. Gilbert RG, Wu AC, Sullivan MA, Sumarriva GE, Ersch N, Hasjim J (2013) Improving human health through understanding the complex structure of glucose polymers. Anal Bioanal Chem 405:8969–8980. https://doi.org/10.1007/s00216-013-7129-1
38. DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, San Carlos
39. The PyMOL molecular graphics system, Version 2.5.2 Schrödinger, LLC
40. Cockburn DW, Cerqueira FM, Bahr C, Koropatkin NM (2020) The structures of the GH13_36 amylases from *Eubacterium rectale* and *Ruminococcus bromii* reveal subsite architectures that favor maltose production. Amylase 4:24–44. https://doi.org/10.1515/amylase-2020-0003
41. Kagawa M, Fujimoto Z, Momma M, Takase K, Mizuno H (2003) Crystal structure of *Bacillus subtilis* alpha-amylase in complex with acarbose. J Bacteriol 185:6981–6984. https://doi.org/10.1128/JB.185.23.6981-6984.2003
42. Abe A, Tonozuka T, Sakano Y, Kamitori S (2004) Complex structures of *Thermoactinomyces vulgaris* R-47 α-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. J Mol Biol 335:811–822. https://doi.org/10.1016/j.jmb.2003.10.078
43. Holm L, Sander C (1995) Dali: a network tool for protein structure comparison. Trends Biochem Sci 20:478–480. https://doi.org/10.1016/s0968-0004(00)89105-7
44. Chai KP, Othman NFB, Teh AH, Ho KL, Chan KG, Shamsir MS, Goh KM, Ng CL (2016) Crystal structure of *Anoxybacillus* α-amylase provides insights into maltose binding of a new glycosyl hydrolase subclass. Sci Rep 6:23126. https://doi.org/10.1038/srep23126
45. Mok SC, Teh AH, Saito JA, Najimudin N, Alam M (2013) Crystal structure of a compact α-amylase from *Geobacillus thermoleovorans*. Enzyme Microb Technol 53:46–54. https://doi.org/10.1016/j.enzmictec.2013.03.009
46. Janeček Š, Kuchtová A, Petrovičová S (2015) A novel GH13 subfamily of α-amylases with a pair of tryptophans in the helix α3 of the catalytic TIM-barrel, the LPDlx signature in the conserved sequence region V and a conserved aromatic motif at the C-terminus. Biologia 70:1284–1294. https://doi.org/10.1515/biolog-2015-0165
47. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z
48. Valguarnera E, Scott NE, Azimzadeh P, Feldman MF (2018) Surface exposure and packing of lipoproteins into outer membrane vesicles are coupled processes in *Bacteroides*. mSphere. 3:e00559-e618. https://doi.org/10.1128/mSphere.00559-18
49. Boel E, Brady L, Brzozowski AM, Derewenda Z, Dodson GG, Jensen VJ, Petersen SB, Swift H, Thim L, Woldike HF (1990) Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1-Å resolution of two enzymes from *Aspergillus*. Biochemistry 29:6244–6249. https://doi.org/10.1021/bi00478a019
50. Kirberger M, Yang JJ (2013) Calcium-binding protein site types. In: Kretsinger RG, Uversky VN, Permyakov EA (eds) Encylcopedia of metalloproteins. Springer, New York, pp 511–521
51. Khrustalev VV, Barkovsky EV, Khrustaleva TA (2016) Magnesium and manganese binding sites on proteins have the same predominant motif of secondary structure. J Theor Biol 395:174–185. https://doi.org/10.1016/j.jtbi.2016.02.006
52. Harding MM (2000) Geometry of metal-ligand interactions in proteins. Acta Crystallogr D Biol Crystallogr 57:401–411. https://doi.org/10.1107/s0907444900019168
53. Gucwa M, Lenkiewicz J, Zheng H, Cymborowski M, Cooper DR, Murzyn K, Minor W (2022) CMM—an enhanced platform for interactive validation of metal binding sites. Protein Sci. https://doi.org/10.1002/pro.4525
54. Nonaka T, Fujihashi M, Kita A, Hagihara H, Ozaki K, Ito S, Miki K (2003) Crystal structure of calcium-free alpha-amylase from *Bacillus* sp. Strain KSM-K38 (AmyK38) and its sodium ion binding sites. J Biol Chem 278:24818–24824. https://doi.org/10.1074/jbc.M212763200
55. Machius M, Declerck N, Huber R, Wiegand G (1998) Activation of *Bacillus licheniformis* alpha-amylase through a disorder–>order transition of the substrate-binding site mediated by a calcium–sodium–calcium metal triad. Structure 6:281–292. https://doi.org/10.1016/s0969-2126(98)00032-x
56. Linden A, Mayans O, Meyer-Klaucke W, Antranikian G, Wilmanns M (2003) Differential regulation of a hyperthermophilic α-amylase with a novel (Ca, Zn) two-metal center by zinc. J Biol Chem. https://doi.org/10.1074/jbc.M211339200
57. Kamitori S, Kondo S, Okuyama K, Yokota T, Shimura Y, Tonozuka T, Sakano Y (1999) Crystal structure of *Thermoactinomyces vulgaris* R-47 alpha-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. J Mol Biol 287:907–921. https://doi.org/10.1006/jmbi.1999.2647
58. Kadziola A, Abe J, Svensson B, Haser R (1994) Crystal and molecular structure of barley alpha-amylase. J Mol Biol 239:104–121. https://doi.org/10.1006/jmbi.1994.1354
59. Hayashi M, Suzuki R, Colleoni C, Ball SG, Fujita N, Suzuki E (2017) Bound substrate in the structure of cyanobacterial branching enzyme supports a new mechanistic model. J Biol Chem 292:5465–5475. https://doi.org/10.1074/jbc.M116.755629
60. Tung JY, Chang MD, Chou WI, Liu YY, Yeh YH, Chang FY, Lin SC, Qiu ZL, Sun YJ (2008) Crystal structures of the starch-binding domain from *Rhizopus oryzae* glucoamylase

reveal a polysaccharide-binding path. Biochem J 416:27–36. https://doi.org/10.1042/BJ20080580
61. Cockburn D, Nielsen MM, Christiansen C, Andersen JM, Rannes JB, Blennow A, Svensson B (2015) Surface binding sites in amylase have distinct roles in recognition of starch structure motifs and degradation. Int J Biol Macromol 75:338–345. https://doi.org/10.1016/j.ijbiomac.2015.01.054
62. Shallom D, Belakhov V, Solomon D, Shoham G, Baasov T, Shoham Y (2002) Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. J Biol Chem 277:43667–43673. https://doi.org/10.1074/jbc.M208285200
63. Park KH, Kim TJ, Cheong TK, Kim JW, Oh BH, Svensson B (2000) Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the alpha-amylase family. Biochim Biophys Acta 1478:165–185. https://doi.org/10.1016/s0167-4838(00)00041-8
64. Kim TJ, Kim MJ, Kim BC, Kim JC, Cheong TK, Kim JW, Park KH (1999) Modes of action of acarbose hydrolysis and transglycosylation catalyzed by a thermostable maltogenic amylase, the gene for which was cloned from a *Thermus* strain. Appl Environ Microbiol 65:1644–1651. https://doi.org/10.1128/AEM.65.4.1644-1651.1999
65. Horváthová V, Janeček Š, Šturdík E (2001) Amylolytic enzymes: molecular aspects of their properties. Gen Physiol Biophys 20:7–32
66. McIver LA, Preuss CV, Tripp J (2022) Acarbose. In: StatPearls. StatPearls Publishing, Treasure Island
67. Ferey-Roux G, Perrier J, Forest E, Marchis-Mouren G, Puigserver A, Santimone M (1998) The human pancreatic alpha-amylase isoforms: isolation, structural studies and kinetics of inhibition by acarbose. Biochim Biophys Acta 1388:10–20. https://doi.org/10.1016/s0167-4838(98)00147-2
68. Lee BH, Eskandari R, Jones K, Reddy KR, Quezada-Calvillo R, Nichols BL, Rose DR, Hamaker BR, Pinto BM (2012) Modulation of starch digestion for slow glucose release through "toggling" of activities of mucosal alpha-glucosidases. J Biol Chem 287:31929–31938. https://doi.org/10.1074/jbc.M112.351858
69. Li C, Begum A, Numao S, Park KH, Withers SG, Brayer GD (2005) Acarbose rearrangement mechanism implied by the kinetic and structural analysis of human pancreatic alpha-amylase in complex with analogues and their elongated counterparts. Biochemistry 44:3347–3357. https://doi.org/10.1021/bi048334e
70. Brzozowski AM, Davies GJ (1997) Structure of the *Aspergillus oryzae* alpha-amylase complexed with the inhibitor acarbose at 2.0 Å resolution. Biochemistry 36:10837–10845. https://doi.org/10.1021/bi970539i
71. Kadziola A, Søgaard M, Svensson B, Haser R (1998) Molecular structure of a barley alpha-amylase-inhibitor complex: implications for starch binding and catalysis. J Mol Biol 278:205–217. https://doi.org/10.1006/jmbi.1998.1683
72. Gilles C, Astier JP, Marchis-Mouren G, Cambillau C, Payan F (1996) Crystal structure of pig pancreatic alpha-amylase isoenzyme II, in complex with the carbohydrate inhibitor acarbose. Eur J Biochem 238:561–569. https://doi.org/10.1111/j.1432-1033.1996.0561z.x
73. Abbott DW, Boraston AB (2012) Quantitative approaches to the analysis of carbohydrate-binding module function. Methods Enzymol 510:211–231. https://doi.org/10.1016/B978-0-12-415931-0.00011-2
74. Cockburn D, Wilkens C, Svensson B (2017) Affinity electrophoresis for analysis of catalytic module-carbohydrate interactions. Methods Mol Biol 1588:119–127. https://doi.org/10.1007/978-1-4939-6899-2_9
75. Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL (2014) Human gut microbes use multiple transporters to distinguish vitamin B12 analogs and compete in the gut. Cell Host Microbe 15:47–57. https://doi.org/10.1016/j.chom.2013.12.007
76. Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, Baslé A, Cartmell A, Terrapon N, Stott K, Lowe EC, McLean R, Shearer K, Schückel J, Venditto I, Ralet MC, Henrissat B, Martens EC, Mosimann SC, Abbott DW, Gilbert HJ (2018) Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic *Bacteroides*. Nat Microbiol 3:210–219. https://doi.org/10.1038/s41564-017-0079-1
77. Bøger M, Hekelaar J, van Leeuwen SS, Dijkhuizen L, Lammerts van Bueren A (2019) Structural and functional characterization of a family GH53 beta-1,4-galactanase from *Bacteroides thetaiotaomicron* that facilitates degradation of prebiotic galactooligosaccharides. J Struct Biol 205:1–10. https://doi.org/10.1016/j.jsb.2018.12.002
78. Kehres DG, Maguire ME (2003) Emerging themes in manganese transport, biochemistry and pathogenesis in bacteria. FEMS Microbiol Rev 27:263–290. https://doi.org/10.1016/S0168-6445(03)00052-4
79. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B (2018) PULDB: the expanded database of polysaccharide utilization loci. Nucleic Acids Res 46:D677–D683. https://doi.org/10.1093/nar/gkx1022
80. Stam MR, Danchin EJG, Rancurel C, Coutinho PM, Henrissat B (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng Des Sel 19:555–562. https://doi.org/10.1093/protein/gzl044
81. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44:W344–W350. https://doi.org/10.1093/nar/gkw408
82. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N (2013) ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr J Chem 53:199–206. https://doi.org/10.1002/ijch.201200096
83. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38:W529–W533. https://doi.org/10.1093/nar/gkq399
84. de Jonge PA, von Meijenfeldt FAB, van Rooijen LE, Brouns SJJ, Dutilh BE (2019) Evolution of BACON domain tandem repeats in crAssphage and novel gut bacteriophage lineages. Viruses. https://doi.org/10.3390/v11121085
85. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klinter S, Pudlo NA, Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens EC, Brumer H (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature 506:498–502. https://doi.org/10.1038/nature12907
86. Lasica AM, Ksiazek M, Madej M, Potempa J (2017) The type IX secretion system (T9SS): highlights and recent insights into its structure and function. Front Cell Infect Microbiol 7:215. https://doi.org/10.3389/fcimb.2017.00215
87. Joglekar P, Sonnenburg ED, Higginbottom SK, Earle KA, Morland C, Shapiro-Ward S, Bolam DN, Sonnenburg JL (2018) Genetic variation of the SusC/SusD homologs from a polysaccharide utilization locus underlies divergent fructan specificities and functional adaptation in *Bacteroides thetaiotaomicron* strains. mSphere. https://doi.org/10.1128/mSphereDirect.00185-18

88. Déjean G, Tamura K, Cabrera A, Jain N, Pudlo NA, Pereira G, Viborg AH, Van Petegem F, Martens EC, Brumer H (2020) Synergy between cell surface glycosidases and glycan-binding proteins dictates the utilization of specific beta(1,3)-glucans by human gut *Bacteroides*. MBio. https://doi.org/10.1128/mBio.00095-20
89. Tamura K, Déjean G, Van Petegem F, Brumer H (2021) Distinct protein architectures mediate species-specific beta-glucan binding and metabolism in the human gut microbiota. J Biol Chem 296:100415. https://doi.org/10.1016/j.jbc.2021.100415
90. Tuncil YE, Xiao Y, Porter NT, Reuhs BL, Martens EC, Hamaker BR (2017) Reciprocal prioritization to dietary glycans by gut bacteria in a competitive environment promotes stable coexistence. MBio 8:e01068-e1117. https://doi.org/10.1128/mBio.01068-17
91. Montanier C, Lammerts van Bueren A, Dumon C, Flint JE, Correia MA, Prates JA, Firbank SJ, Lewis RJ, Grondin GG, Ghinet MG, Gloster TM, Herve C, Knox JP, Talbot BG, Turkenburg JP, Kerovuo J, Brzezinski R, Fontes CMGA, Davies GJ, Boraston AB, Gilbert HJ (2009) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. Proc Natl Acad Sci U S A 106:3065–3070. https://doi.org/10.1073/pnas.0808972106
92. Ezer A, Matalon E, Jindou S, Borovok I, Atamna N, Yu Z, Morrison M, Bayer EA, Lamed R (2008) Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. J Bacteriol 190:8220–8222. https://doi.org/10.1128/JB.00609-08
93. Møller MS, Henriksen A, Svensson B (2016) Structure and function of alpha-glucan debranching enzymes. Cell Mol Life Sci 73:2619–2641. https://doi.org/10.1007/s00018-016-2241-y
94. Feng L, Fawaz R, Hovde S, Gilbert L, Chiou J, Geiger JH (2015) Crystal structures of *Escherichia coli* branching enzyme in complex with linear oligosaccharides. Biochemistry 54:6207–6218. https://doi.org/10.1021/acs.biochem.5b00228
95. Janeček Š, Svensson B, MacGregor EA (2011) Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzyme Microb Technol 49:429–440. https://doi.org/10.1016/j.enzmictec.2011.07.002
96. White JBR, Silale A, Feasey M, Heunis T, Zhu Y, Zheng H, Gajbhiye A, Firbank S, Baslé A, Trost M, Bolam DN, van den Berg B, Ranson NA (2023) Outer membrane utilisomes mediate glycan uptake in gut Bacteroidetes. Nature. 618:583–589. https://doi.org/10.1038/s41586-023-06146-w
97. Thompson J, Hess S, Pikis A (2004) Genes *malh* and *pagl* of *Clostridium acetobutylicum* ATCC 824 encode NAD^+- and Mn^{2+}-dependent phospho-alpha-glucosidase(s). J Biol Chem 279:1553–1561. https://doi.org/10.1074/jbc.M310733200
98. Raasch C, Streit W, Schanzer J, Bibel M, Gosslar U, Liebl W (2000) *Thermotoga maritima* AglA, an extremely thermostable NAD^+-, Mn^{2+}-, and thiol-dependent alpha-glucosidase. Extremophiles 4:189–200. https://doi.org/10.1007/pl00010711
99. Nielsen FH (2012) Manganese, molybdenum, boron, chromium, and other trace elements. In: Erdman JW Jr, Macdonald IA, Zeisel SH (eds) Present knowledge in nutrition, 10th edn. Wiley-Blackwell, Ames, pp 586–607
100. Guiberson ER, Good CJ, Wexler AG, Skaar EP, Spraggins JM, Caprioli RM (2022) Multimodal imaging mass spectrometry of murine gastrointestinal tract with retained luminal content. J Am Soc Mass Spectrom 33:1073–1076. https://doi.org/10.1021/jasms.1c00360
101. Davis CD, Zech L, Greger JL (1993) Manganese metabolism in rats: an improved methodology for assessing gut endogenous losses. Proc Soc Exp Biol Med 202:103–108. https://doi.org/10.3181/00379727-202-43518
102. Yu X, Han J, Li H, Zhang Y, Feng J (2018) The effect of enzymes on release of trace elements in feedstuffs based on in vitro digestion model for monogastric livestock. J Anim Sci Biotechnol 9:73. https://doi.org/10.1186/s40104-018-0289-2
103. Pajarillo EAB, Lee E, Kang DK (2021) Trace metals and animal health: interplay of the gut microbiota with iron, manganese, zinc, and copper. Anim Nutr 7:750–761. https://doi.org/10.1016/j.aninu.2021.03.005
104. Lopez CA, Skaar EP (2018) The impact of dietary transition metals on host-bacterial interactions. Cell Host Microbe 23:737–748. https://doi.org/10.1016/j.chom.2018.05.008
105. Damo S, Chazin WJ, Skaar EP, Kehl-Fie TE (2012) Inhibition of bacterial superoxide defense: a new front in the struggle between host and pathogen. Virulence 3:325–328. https://doi.org/10.4161/viru.19635
106. Diaz-Ochoa VE, Lam D, Lee CS, Klaus S, Behnsen J, Liu JZ, Chim N, Nuccio SP, Rathi SG, Mastroianni JR, Edwards RA, Jacobo CM, Cerasi M, Battistoni A, Ouellette AJ, Goulding CW, Chazin WJ, Skaar EP, Raffatellu M (2016) *Salmonella* mitigates oxidative stress and thrives in the inflamed gut by evading calprotectin-mediated manganese sequestration. Cell Host Microbe 19:814–825. https://doi.org/10.1016/j.chom.2016.05.005
107. DeMartino P, Cockburn DW (2020) Resistant starch: impact on the gut microbiome and health. Curr Opin Biotechnol 61:66–71. https://doi.org/10.1016/j.copbio.2019.10.008
108. Cerqueira FM, Photenhauer AL, Pollet RM, Brown HA, Koropatkin NM (2020) Starch digestion by gut bacteria: crowdsourcing for carbs. Trends Microbiol 28:95–108. https://doi.org/10.1016/j.tim.2019.09.004
109. Gu Y, Wang X, Li J, Zhang Y, Zhong H, Liu R, Zhang D, Feng Q, Xie X, Hong J, Ren H, Liu W, Ma J, Su Q, Zhang H, Yang J, Wang X, Zhao X, Gu W, Bi Y, Peng Y, Xu X, Xia H, Li F, Xu X, Yang H, Xu G, Madsen L, Kristiansen K, Ning G, Wang W (2017) Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. Nat Commun 8:1785. https://doi.org/10.1038/s41467-017-01682-2
110. Pudlo NA, Urs K, Crawford R, Pirani A, Atherly T, Jimenez R, Terrapon N, Henrissat B, Peterson D, Ziemer C, Snitkin E, Martens EC (2022) Phenotypic and genomic diversification in complex carbohydrate-degrading human gut bacteria. mSystems. 7:e0094721. https://doi.org/10.1128/msystems.00947-21
111. Holdeman LV, Moore WEC (eds) (1977) Anaerobe laboratory manual, 4th edn. Virginia Polytechnic Institute and State University, Blacksburg
112. Koropatkin NM, Martens EC, Gordon JI, Smith TJ (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16:1105–1115. https://doi.org/10.1016/j.str.2008.03.017
113. Winter G, Lobley CM, Prince SM (2013) Decision making in *xia*2. Acta Crystallogr D Biol Crystallogr 69:1260–1273. https://doi.org/10.1107/S0907444913015308
114. Kabsch W (2010) XDS. Acta Crystallogr D Biol Crystallogr 66:125–132. https://doi.org/10.1107/S0907444909047337
115. Long F, Vagin AA, Young P, Murshudov GN (2008) *BALBES*: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr 64:125–132. https://doi.org/10.1107/S0907444907050172
116. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD (2008) Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. Acta Crystallogr D Biol Crystallogr 64:61–69. https://doi.org/10.1107/S090744490705024X

117. Emsley P, Cowtan K (2004) *Coot*: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132. https://doi.org/10.1107/S0907444904019158
118. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Žídek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D, Velankar S (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50:D439–D444. https://doi.org/10.1093/nar/gkab1061
119. Zhai X, Wu K, Ji R, Zhao Y, Lu J, Yu Z, Xu X, Huang J (2022) Structure and function insight of the alpha-glucosidase QsGH13 from *Qipengyuania seohaensis* sp. Sw-135. Front Microbiol 13:849585. https://doi.org/10.3389/fmicb.2022.849585
120. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493
121. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) *REFMAC5* for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367. https://doi.org/10.1107/S0907444911001314
122. Beilsten-Edmands J, Winter G, Gildea R, Parkhurst J, Waterman D, Evans G (2020) Scaling diffraction data in the *DIALS* software package: algorithms and new approaches for multi-crystal scaling. Acta Crystallogr D Struct Biol 76:385–399. https://doi.org/10.1107/S2059798320003198
123. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) *Phaser* crystallographic software. J Appl Crystallogr 40:658–674. https://doi.org/10.1107/S0021889807021206
124. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Crystallogr D Struct Biol 75:861–877. https://doi.org/10.1107/S2059798319011471
125. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) *PHENIX*: a comprehensive python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221. https://doi.org/10.1107/S0907444909052925
126. Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69:1204–1214. https://doi.org/10.1107/S0907444913000061
127. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) *PHENIX*: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58:1948–1954. https://doi.org/10.1107/s0907444902016657
128. Agirre J, Iglesias-Fernández J, Rovira C, Davies GJ, Wilson KS, Cowtan KD (2015) Privateer: software for the conformational validation of carbohydrate structures. Nat Struct Mol Biol 22:833–834. https://doi.org/10.1038/nsmb.3115
129. Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428. https://doi.org/10.1021/ac60147a030
130. McKee LS (2017) Measuring enzyme kinetics of glycoside hydrolases using the 3,5-dinitrosalicylic acid assay. Methods Mol Biol 1588:27–36. https://doi.org/10.1007/978-1-4939-6899-2_3
131. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75
132. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 50:W276–W279. https://doi.org/10.1093/nar/gkac240
133. Cameron EA, Maynard MA, Smith CJ, Smith TJ, Koropatkin NM, Martens EC (2012) Multidomain carbohydrate-binding proteins involved in *Bacteroides thetaiotaomicron* starch metabolism. J Biol Chem 287:34614–34625. https://doi.org/10.1074/jbc.M112.397380
134. Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453. https://doi.org/10.1016/0022-2836(70)90057-4
135. DNASTAR, MegAlign Pro ®. Version 17.3.058. Madison, WI
136. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, Gascuel O (2019) NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res 47:W260–W265. https://doi.org/10.1093/nar/gkz303
137. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
138. Criscuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10:210. https://doi.org/10.1186/1471-2148-10-210
139. Lefort V, Desper R, Gascuel O (2015) FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol 32:2798–2800. https://doi.org/10.1093/molbev/msv150
140. Junier T, Zdobnov EM (2010) The Newick utilities: high-throughput phylogenetic tree processing in the unix shell. Bioinformatics 26:1669–1670. https://doi.org/10.1093/bioinformatics/btq243
141. Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296. https://doi.org/10.1093/nar/gkab301
142. Sorimachi K, Le Gal-Coëffet MF, Williamson G, Archer DB, Williamson MP (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to beta-cyclodextrin. Structure 5:647–661. https://doi.org/10.1016/s0969-2126(97)00220-7
143. Møller MS, Abou Hachem M, Svensson B, Henriksen A (2012) Structure of the starch-debranching enzyme barley limit dextrinase reveals homology of the N-terminal domain to CBM21. Acta Crystallogr Sect F Struct Biol Cryst Commun 68:1008–1012. https://doi.org/10.1107/S1744309112031004
144. Marty MT, Baldwin AJ, Marklund EG, Hochberg GK, Benesch JL, Robinson CV (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem 87:4370–4376. https://doi.org/10.1021/acs.analchem.5b00140
145. Mareček F, Janeček Š (2022) A novel subfamily GH13_46 of the α-amylase family GH13 represented by the cyclomaltodextrinase

from *Flavobacterium* sp. No 92. Molecules 27:8735. https://doi.org/10.3390/molecules27248735